

Measurement uncertainty relations for discrete observables: Relative entropy formulation

Alberto Barchielli^{1,2,3}, Matteo Gregoratti^{1,2}, Alessandro Toigo^{1,3}

¹ Politecnico di Milano, Dipartimento di Matematica,
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy,

² Istituto Nazionale di Alta Matematica (INDAM-GNAMPA),

³ Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Milano

8th August 2016

Abstract

We introduce a new informational-theoretic formulation of the quantum measurement uncertainty relations, based on the notion of relative entropy between measurement probabilities. In the case of a finite-dimensional system, we quantify the total error affecting an approximate joint measurement of two discrete observables, we prove the general properties of its minimum value (the uncertainty lower bound) and we study the corresponding optimal approximate joint measurements. The new error bound, which we name *entropic incompatibility degree*, turns out to enjoy many key features: among the main ones, it is state independent and tight, it shares the desirable invariance properties, and it vanishes if and only if the two observables are compatible. In this context, we point out the difference between generic approximate joint measurements and sequential approximate joint measurements; to do this, we introduce a separate index for the tradeoff between the error in the first measurement and the disturbance of the second one. By exploiting the symmetry properties of the target observables, exact values and lower bounds are computed in two different concrete examples: (1) a couple of spin-1/2 components (not necessarily orthogonal); (2) two Fourier conjugate mutually unbiased bases in prime power dimension. Finally, the measurement uncertainty relations are generalized to the case of many observables.

1 Introduction

In the foundations of Quantum Mechanics, a remarkable achievement of the last years has been the clarification of the differences between *preparation uncertainty relations* (PUR) and *measurement uncertainty relations* (MUR) [1–11], both of them arising from the Heisenberg’s heuristic considerations about the precision with which the position and the momentum of a quantum particle can be determined [12].

One speaks of PUR when some lower bound is given on the “spreads” of the distributions of two observables A and B measured in the same state ρ . The most known formulation of PUR, due to Robertson [13], involves the product of the two standard deviations; more recent formulations are given in terms of distances among probability distributions [11] or entropies [14–18].

On the other hand, one refers to MUR when some lower bound is given on the “errors” of any approximate joint measurement M of two target observables A and B . When the approximate joint measurement M is realized as a sequence of two measurements, one for each target observable, MUR are regarded also as relations between the “error” allowed in an approximate measurement of the first observable and the “disturbance” affecting the successive measurement of the second one.

Although the recent developments of the theory of approximate quantum measurements [19–22] and nondisturbing quantum measurements [23, 24] have generated a considerable renewed interest in MUR, no agreement has yet been reached about the proper quantifications of the “error” or “disturbance” terms. Here, the main problem is how to compare the target observables A and B with their approximate or

perturbed versions provided by the marginals $M_{[1]}$ and $M_{[2]}$ of M ; indeed, A , $M_{[1]}$, $M_{[2]}$ and B may typically be incompatible. The proposals then range from operator formulations of the error [1–4, 25, 26] to distances for probability distributions [5–7, 9, 11] and conditional entropies [27–29].

In this paper, we propose and develop a new approach to MUR based on the notion of *relative entropy*. Here we will deal with the case of discrete observables for a finite dimensional quantum system.

In the spirit of [5–7, 11], we quantify the “error” in the approximation $A \simeq M_{[1]}$ by comparing the respective outcome distributions A^ρ and $M_{[1]}^\rho$ in every possible state ρ ; however, differently from [5–7, 11], the comparison is done from the point of view of information theory. Then, the natural choice is to consider $S(A^\rho \| M_{[1]}^\rho)$, the relative entropy of A^ρ with respect to $M_{[1]}^\rho$, as a quantification of the information loss when A^ρ is approximated with $M_{[1]}^\rho$. Similarly, in order to quantify either the “error” or – if A and B are measured in sequence – the “disturbance” related to the approximation $B \simeq M_{[2]}$, we employ the relative entropy $S(B^\rho \| M_{[2]}^\rho)$. Relative entropy appears to be the fundamental quantity from which the other entropic notions can be derived, cf. [30, 31], [32, p. 15]; we recall its relevant properties in Section 1.2. It should be noticed that the relative entropy, of classical or quantum type, has already been used in quantum measurement theory to give proper measures of information gains and losses in various scenarios [32–36].

The main difference between our relative entropy approach and the conditional entropy approach of [27–29] is that we do not need any joint distribution for A^ρ , $M_{[1]}^\rho$, $M_{[2]}^\rho$ and B^ρ , which indeed would not exist for incompatible observables.

The relative entropy formulation of MUR is: for every approximate joint measurement M of A and B , there exists a state ρ such that

$$S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \geq c(A, B),$$

where the uncertainty lower bound

$$c(A, B) = \inf_M \sup_\rho \left\{ S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \right\}$$

depends on the allowed joint measurements M . We stress that, in the above definition, the same state ρ appears in both the error terms $S(A^\rho \| M_{[1]}^\rho)$ and $S(B^\rho \| M_{[2]}^\rho)$; thus, all possible error compensations are taken into account in the maximization. This is essentially different from the approach of [5–7, 11], where instead the two errors are maximized in separate stages.

By considering any possible approximate joint measurement in the definition of $c(A, B)$, we get an uncertainty lower bound $c_{\text{inc}}(A, B)$ that turns out to be a proper measure of the incompatibility of A and B . On the other hand, by considering only sequential measurements, we derive an uncertainty lower bound $c_{\text{ed}}(A, B)$ that provides a suitable quantification of the error/disturbance tradeoff for the two (sequentially ordered) target observables. Indeed, such lower bounds share a lot of desirable properties: they are zero if and only if the target observables are compatible (respectively, sequentially compatible); they are invariant under unitary transformations and relabelling of the output values of the measurements; and finally they are bounded from above by a value that is independent of both the dimension of the Hilbert space and the number of the possible outcomes. As a main result, we show also that, for a generic couple of observables A and B , considering only their sequential measurements is a real restriction, because in general $c_{\text{ed}}(A, B)$ may be larger than $c_{\text{inc}}(A, B)$; actually, the two indexes are guaranteed to coincide only if one makes some extra assumptions on A and B , for example, if the second observable B is supposed to be sharp.

Thus, every time A and B are incompatible, the total loss of information $S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho)$ in the approximations $A \simeq M_{[1]}$ and $B \simeq M_{[2]}$ depends on both the joint measurement M and the state ρ ; however, since $c_{\text{inc}}(A, B) > 0$, we prove that there is a minimum potential loss that no joint measurement M can avoid. Similar remarks hold for sequential joint measurements and the corresponding error/disturbance tradeoff. Whenever A and B are incompatible, we will look for the exact value of $c_{\text{inc}}(A, B)$, or at least some lower bound for it, as well as we will try to determine the optimal approximate joint measurements M which saturate the minimum. In particular, we will prove that in some relevant applications there is actually a unique such M , thus showing that in these cases the entropic optimality criterium unambiguously fixes the best approximate joint measurement.

The generalization of the relative entropy formulation of MUR to the case of more than two target observables is rather straightforward, and it will be illustrated at the end of the paper. It is worth noticing

that there are triples of observables whose optimal approximate joint measurements are not unique, even if all their possible pairings do have the corresponding binary uniqueness property (see e.g. the two and three orthogonal spin-1/2 components in Section 3.2 and Example 2).

Now, we briefly summarize the content of the paper. In Section 2, we introduce and study the entropic divergence $D(A, B \| M)$ of an arbitrary approximate joint measurement M of A and B from its two target observables, that is, the sum $S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho)$ evaluated at the worst possible system state ρ . This gives the state-independent quantification of the total “error” made with the approximate joint measurement M at hand. Then we find the uncertainty lower bounds $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ by considering the best approximate joint measurement M and we prove their general properties. In Section 3, we undertake the explicit computation of $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ and their respective optimal approximate joint measurements M for several examples of incompatible target observables. Some general results are proved, which show how the symmetry properties of the quantum system can help in the task. Then, two cases are studied: two spin-1/2 components, which we do not assume to be necessarily orthogonal, and two Fourier conjugate observables associated with a pair of mutually unbiased bases (MUB) in prime power dimension. Finally, in Section 4 we generalize the relative entropy formulation of MUR to the case of many target observables. As an example, the case of three orthogonal spin-1/2 components is completely solved.

1.1 Observables and instruments

We start by fixing our quantum system and recalling the notions and basic facts on observables and measurements that we will use in the article [19, 20, 22, 37–39].

The Hilbert space \mathcal{H} and the spaces $\mathcal{L}(\mathcal{H})$, $\mathcal{T}(\mathcal{H})$, $\mathcal{S}(\mathcal{H})$. We consider a quantum system described by a finite-dimensional complex Hilbert space \mathcal{H} , with $\dim \mathcal{H} = d$; then, the spaces $\mathcal{L}(\mathcal{H})$ of the linear bounded operators on \mathcal{H} and the trace-class $\mathcal{T}(\mathcal{H})$ coincide. By $\mathcal{S}(\mathcal{H})$ we denote the convex set of the states on \mathcal{H} (positive, unit trace operators), which is a compact subset of $\mathcal{T}(\mathcal{H})$. The extreme points of $\mathcal{S}(\mathcal{H})$ are the pure states (rank-one projections) $\rho = |\psi\rangle\langle\psi|$, with $\psi \in \mathcal{H}$ and $\|\psi\| = 1$.

The spaces of observables $\mathcal{M}(\mathcal{X})$ and of probabilities $\mathcal{P}(\mathcal{X})$. In the general formulation of quantum mechanics, an *observable* is identified with a *positive operator valued measure* (POVM). We will consider only observables with outcomes in a finite set \mathcal{X} (with its power set as its σ -algebra). Then, a POVM on \mathcal{X} is identified with its discrete density $A : \mathcal{X} \rightarrow \mathcal{L}(\mathcal{H})$, whose values $A(x)$ are positive operators on \mathcal{H} such that $\sum_{x \in \mathcal{X}} A(x) = \mathbb{1}$; here, the sum has a finite number $|\mathcal{X}|$ of terms ($|\mathcal{X}|$ denotes the cardinality of \mathcal{X}). Similarly, a probability on \mathcal{X} is identified with its discrete probability density (or mass function) $p : \mathcal{X} \rightarrow \mathbb{R}$, where $p(x) \geq 0$ and $\sum_{x \in \mathcal{X}} p(x) = 1$.

For $\rho \in \mathcal{S}(\mathcal{H})$, the function $A^\rho(x) = \text{Tr}\{\rho A(x)\}$ is the discrete probability density on \mathcal{X} which gives the outcome distribution in a measurement of the observable A performed on the quantum system prepared in the state ρ .

We denote by $\mathcal{M}(\mathcal{X})$ the set of observables which are associated with the system at hand and have outcomes in \mathcal{X} ; $\mathcal{M}(\mathcal{X})$ is a convex, compact subset of $\mathcal{L}(\mathcal{H})^{\mathcal{X}}$, the finite dimensional linear space of the functions from \mathcal{X} to $\mathcal{L}(\mathcal{H})$. Both the mappings $\rho \mapsto A^\rho$ and $A \mapsto A^\rho$ are continuous and affine (i.e. preserving convex combinations) from the respective domains into the convex set $\mathcal{P}(\mathcal{X})$ of the probabilities on \mathcal{X} . As a subset of $\mathbb{R}^{\mathcal{X}}$, the set $\mathcal{P}(\mathcal{X})$ is convex and compact. The extreme points of $\mathcal{P}(\mathcal{X})$ are the delta distributions (Kronecker deltas) δ_x , with $x \in \mathcal{X}$.

Trivial and sharp observables. An observable A is *trivial* if $A = p\mathbb{1}$ for some probability p , where $\mathbb{1}$ is the identity of \mathcal{H} . In particular, we will make use of the uniform distribution $u_{\mathcal{X}}$ on \mathcal{X} , $u_{\mathcal{X}}(x) = 1/|\mathcal{X}|$, and the trivial uniform observable $U_{\mathcal{X}} = u_{\mathcal{X}}\mathbb{1}$.

An observable A is *sharp* if $A(x)$ is a projection $\forall x \in \mathcal{X}$. Note that we allow $A(x) = 0$ for some x , which is required when dealing with sets of observables sharing the same outcome space. Of course, for every sharp observable we have $|\{x : A(x) \neq 0\}| \leq d$.

Bi-observables and compatible observables. When the outcome set has the product form $\mathcal{X} \times \mathcal{Y}$, we speak of bi-observables. In this case, given the POVM $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we can introduce also the two *marginal observables* $M_{[1]} \in \mathcal{M}(\mathcal{X})$ and $M_{[2]} \in \mathcal{M}(\mathcal{Y})$ by

$$M_{[1]}(x) = \sum_{y \in \mathcal{Y}} M(x, y), \quad M_{[2]}(y) = \sum_{x \in \mathcal{X}} M(x, y).$$

In the same way, for $p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we get the definition of the *marginal probabilities* $p_{[1]} \in \mathcal{P}(\mathcal{X})$ and $p_{[2]} \in \mathcal{P}(\mathcal{Y})$. Clearly, $(M_{[i]})^\rho = (M^\rho)_{[i]}$; hence there is no ambiguity in writing $M_{[i]}^\rho$ for both the probabilities.

On the other hand, we say that two observables $A \in \mathcal{M}(\mathcal{X})$ and $B \in \mathcal{M}(\mathcal{Y})$ are *jointly measurable* or *compatible* if there exists a bi-observable $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ such that $M_{[1]} = A$ and $M_{[2]} = B$; we then call M a *joint measurement* of A and B .

Two classical probabilities $p \in \mathcal{P}(\mathcal{X})$ and $q \in \mathcal{P}(\mathcal{Y})$ are always compatible, as they can be seen as the marginals of at least one joint probability in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Indeed, one can take the *product probability* $p \otimes q$ given by $(p \otimes q)(x, y) = p(x)q(y)$. Clearly, nothing similar can be defined for two arbitrary quantum observables, for which instead compatibility usually is a highly nontrivial requirement.

The space of instruments $\mathcal{J}(\mathcal{X})$. Given a pre-measurement state ρ , a POVM allows to compute the probability distribution of the measurement outcome. In order to describe also the change of the state produced by the measurement, we need the more general mathematical notion of an *instrument*, i.e. a measure \mathcal{J} on the outcome set \mathcal{X} taking values in the set of the completely positive maps on $\mathcal{L}(\mathcal{H})$. In our case of finitely many outcomes, an instrument is described by its discrete density $x \mapsto \mathcal{J}_x$, $x \in \mathcal{X}$, whose general structure is $\mathcal{J}_x[\rho] = \sum_{\alpha} J_x^\alpha \rho J_x^{\alpha*}$, $\forall \rho \in \mathcal{S}(\mathcal{H})$; here, the Kraus operators $J_x^\alpha \in \mathcal{L}(\mathcal{H})$ are such that $\sum_{x \in \mathcal{X}} \sum_{\alpha} J_x^{\alpha*} J_x^\alpha = \mathbb{1}$ and, since \mathcal{H} is finite-dimensional, the index α can be restricted to finitely many values. The *adjoint instrument* is given by $\mathcal{J}_x^*[F] = \sum_{\alpha} J_x^{\alpha*} F J_x^\alpha$, $\forall F \in \mathcal{L}(\mathcal{H})$. The sum $\mathcal{J}_{\mathcal{X}} := \sum_{x \in \mathcal{X}} \mathcal{J}_x$ is a *quantum channel*, i.e. a completely positive trace preserving map on $\mathcal{S}(\mathcal{H})$. We denote by $\mathcal{J}(\mathcal{X})$ the convex and compact set of the \mathcal{X} -valued instruments for our quantum system.

By setting $A(x) := \mathcal{J}_x^*[\mathbb{1}] = \sum_{\alpha} J_x^{\alpha*} J_x^\alpha$, a POVM $A \in \mathcal{M}(\mathcal{X})$ is defined, which is the observable measured by the instrument \mathcal{J} ; we say that *the instrument \mathcal{J} implements the observable A* . The state of the system after the measurement, conditioned on the outcome x , is $\mathcal{J}_x[\rho]/A^\rho(x)$. We recall that, given an observable A , one can always find an instrument \mathcal{J} implementing A , but \mathcal{J} is not uniquely determined by A , i.e. different instruments \mathcal{J} , with different actions on the quantum system, may be used to measure the same observable A .

Sequential measurements and sequentially compatible observables. Employing the notion of instrument, we can describe a measurement of an observables $A \in \mathcal{M}(\mathcal{X})$ followed by a measurement of an observable $B \in \mathcal{M}(\mathcal{Y})$: a *sequential measurement* of A followed by B is a bi-observable $M(x, y) = \mathcal{J}_x^*[B(y)]$, where \mathcal{J} is any instrument implementing A . Its marginals are $M_{[1]}(x) = \mathcal{J}_x^*[\mathbb{1}] = A(x)$ and $M_{[2]}(y) = \mathcal{J}_{\mathcal{X}}^*[B(y)]$. We write $M = \mathcal{J}^*(B)$, which is a measurement in which one first applies the instrument \mathcal{J} to measure A , and then he measures the observable B on the resulting output state; in this way, he obtains a joint measurement of A and $\mathcal{J}_{\mathcal{X}}^*[B(\cdot)]$, a perturbed version of B .

An observable $A \in \mathcal{M}(\mathcal{X})$ can be measured without disturbing $B \in \mathcal{M}(\mathcal{Y})$ [23], or shortly A and B are *sequentially compatible observables*, if there exists a sequential measurement $M = \mathcal{J}^*(B)$ such that

$$M_{[1]} \equiv \mathcal{J}^*[\mathbb{1}] = A, \quad M_{[2]} \equiv \mathcal{J}_{\mathcal{X}}^*[B(\cdot)] = B.$$

In this case, a measurement of B at time 1 (i.e. after the measurement of A) has the same outcome distribution of a measurement of B at time 0 (i.e. before the measurement of A).

If A and B are sequentially compatible observables, then clearly they are also jointly measurable. However, the opposite is not true; two counterexample are shown in [23] and are reported in Appendix A. This happens because we demand to measure just B at time 1, i.e. we do not content ourselves with getting at time 1 the same outcome distribution of a measurement of B performed at time 0. Indeed, this second requirement is weaker: it can be satisfied by any couple of jointly measurable observables A and B , by

measuring a suitable observable C after A (with A implemented by an instrument \mathcal{I} which possibly increases the dimension of the Hilbert space). The definition of sequentially compatible observables is not symmetric, and indeed there exist couples of observables such that A can be measured without disturbing B , but for which the opposite is not true. This asymmetry also reflects in the remarkable fact that, if the second observable is sharp, then the compatibility of A and B turns out to be equivalent to their sequential compatibility.

Reference observables. In this paper, we will fix two target observables with finitely many values, $A \in \mathcal{M}(\mathcal{X})$ and $B \in \mathcal{M}(\mathcal{Y})$, and we will study several ways of characterizing their uncertainty relations. For any $\rho \in \mathcal{S}(\mathcal{H})$, the associated probability distributions A^ρ and B^ρ can be estimated by measuring either A or B in many identical preparations of the quantum system in the state ρ . No joint or sequential measurement of A and B is required at this stage. In Section 2 we will develop a general theory to quantify the error made by approximating A and B with compatible observables and introduce the notion of optimal approximate joint measurement for them.

1.2 Shannon and relative entropies

In this paper, we will be concerned with entropic quantities of classical type [30, 31]; we will express them in “bits”, which means to use logarithms with base 2: $\log \equiv \log_2$.

The fundamental quantity is the *relative entropy*, or *information divergence*. Given two probabilities $p, q \in \mathcal{P}(\mathcal{X})$, the relative entropy of p with respect to q (also called the divergence of q from p) is

$$S(p||q) = \begin{cases} \sum_{x \in \text{supp } p} p(x) \log \frac{p(x)}{q(x)} & \text{if } \text{supp } p \subseteq \text{supp } q, \\ +\infty & \text{otherwise,} \end{cases} \quad (1)$$

which defines an extended real function on the product set $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. Also the terms *Kullback-Leibler divergence* and *information for discrimination* are used for $S(p||q)$. Instead, the *Shannon entropy* of the probability p is defined by

$$H(p) = - \sum_{x \in \mathcal{X}} p(x) \log p(x).$$

The Shannon entropy of p is a measure of the uncertainty of a random variable with distribution p [31, Sect. 2.1]. On the other hand, the relative entropy $S(p||q)$ is a measure of the inefficiency of assuming that the probability is q when the true probability is p [31, Sect. 2.3]; in other words, it is the amount of information lost when q is used to approximate p [30, p. 51]. Let us stress that $S(p||q)$ compares p and q , but it is not a distance since it is not symmetric. As such, the use of S is particularly convenient when the two probabilities have different roles; for instance, if p is the true distribution of a given random variable, while q is the distribution actually used as an approximation of p . This will be our case, where the role of p is played by the distribution A^ρ (or B^ρ) of the target observable A (or B) and q will be the distribution of some allowed approximation. Note, in particular, that no joint distribution of p and q is involved.

Let us collect in the following proposition the main properties of the Shannon and relative entropies [19, 31, 32, 38, 40]. For the definition and main properties of lower semicontinuous (LSC) functions, we refer to [41, Sect. 1.5].

Proposition 1. *The following properties hold:*

- (i) $0 \leq H(p) \leq \log |\mathcal{X}|$ and $S(p||q) \geq 0$, for all $p, q \in \mathcal{P}(\mathcal{X})$.
- (ii) $H(p) = 0$ if and only if $p = \delta_x$ for some x , where δ_x is the delta distribution at x . $S(p||q) = 0$ if and only if $p = q$.
- (iii) $H(u_{\mathcal{X}}) = \log |\mathcal{X}|$, where $u_{\mathcal{X}}$ is the uniform probability on \mathcal{X} , and $S(p||u_{\mathcal{X}}) = \log |\mathcal{X}| - H(p)$ for all $p \in \mathcal{P}(\mathcal{X})$.

(iv) H and S are invariant for relabelling of the outcomes; that is, if $f : \mathcal{X}' \rightarrow \mathcal{X}$ is a bijective map, then $H(p \circ f) = H(p)$ and $S(p \circ f \| q \circ f) = S(p \| q)$.

(v) H is a concave function on $\mathcal{P}(\mathcal{X})$, and S is jointly convex on $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$, namely

$$S(\lambda p_1 + (1 - \lambda)p_2 \| \lambda q_1 + (1 - \lambda)q_2) \leq \lambda S(p_1 \| q_1) + (1 - \lambda)S(p_2 \| q_2), \quad \forall \lambda \in [0, 1].$$

(vi) The function $p \mapsto H(p)$ is continuous on $\mathcal{P}(\mathcal{X})$. The function $(p, q) \mapsto S(p \| q)$ is LSC on $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$.

(vii) If $p_1, p_2 \in \mathcal{P}(\mathcal{X})$ and $q_1, q_2 \in \mathcal{P}(\mathcal{Y})$, then $S(p_1 \otimes q_1 \| p_2 \otimes q_2) = S(p_1 \| p_2) + S(q_1 \| q_2)$.

In order to study some further specific properties of the relative entropy that will be needed in the following, it is useful to introduce the extended real function $s : [0, 1] \times [0, 1] \rightarrow [-1/2, +\infty]$, with

$$s(u, v) = \begin{cases} u \log \frac{u}{v} & \text{if } 0 < u \leq 1 \text{ and } 0 < v \leq 1, \\ 0 & \text{if } u = 0 \text{ and } 0 \leq v \leq 1, \\ +\infty & \text{if } u > 0 \text{ and } v = 0. \end{cases} \quad (2)$$

In terms of the function s , the relative entropy can be rewritten as

$$S(p \| q) = \sum_{x \in \mathcal{X}} s(p(x), q(x)).$$

Note that, unlike the relative entropy, s can take also negative values, and its minimum is $s(1/2, 1) = -1/2$. As a function of (u, v) , s is continuous at all the points of the square $[0, 1] \times [0, 1]$ except at the origin $(0, 0)$, where it can be easily proved that it is only LSC.

We collect in the next proposition some more facts about S .

Proposition 2. *The following properties hold:*

- (i) For all $p, q_1, q_2 \in \mathcal{P}(\mathcal{X})$, the extended real function $f(\lambda) = S(p \| \lambda q_1 + (1 - \lambda)q_2)$ is convex in $\lambda \in [a, b]$, where $a \leq 0$ and $b \geq 1$ are the minimum and maximum values of λ , respectively, such that $\lambda q_1 + (1 - \lambda)q_2$ is a probability distribution on \mathcal{X} .
- (ii) For all $\lambda \in (0, 1]$ and $q \in \mathcal{P}(\mathcal{X})$, the function $g_\lambda(p) = S(p \| \lambda p + (1 - \lambda)q)$ is finite and continuous in $p \in \mathcal{P}(\mathcal{X})$. It attains the maximum value

$$\max_{p \in \mathcal{P}(\mathcal{X})} S(p \| \lambda p + (1 - \lambda)q) = \log \frac{1}{\lambda + (1 - \lambda) \min_{x \in \mathcal{X}} q(x)}, \quad (3)$$

which is a strictly decreasing function of $\lambda \in (0, 1]$.

Proof. (i) The mapping $\lambda \mapsto \lambda q_1 + (1 - \lambda)q_2$ is affine on the interval $[a, b]$, and the relative entropy is convex in its second entry by item (v) of Proposition 1. It follows that the composition $\lambda \mapsto S(p \| \lambda q_1 + (1 - \lambda)q_2)$ is convex on $[a, b]$.

(ii) Let $\lambda \in (0, 1]$. For all $u, v \in [0, 1]$, the condition $u > 0$ implies that $\lambda u + (1 - \lambda)v > 0$, hence

$$s(u, \lambda u + (1 - \lambda)v) = \begin{cases} u \log \frac{u}{\lambda u + (1 - \lambda)v} & \text{if } 0 < u \leq 1, \\ 0 & \text{if } u = 0. \end{cases}$$

Clearly, this is a continuous function of $u \in (0, 1]$. To see that it is continuous also at 0, we take the limit

$$\begin{aligned} \lim_{u \rightarrow 0^+} u \log \frac{u}{\lambda u + (1 - \lambda)v} &= \lim_{u \rightarrow 0^+} u \log u - \lim_{u \rightarrow 0^+} u \log [\lambda u + (1 - \lambda)v] \\ &= - \lim_{u \rightarrow 0^+} u \log [\lambda u + (1 - \lambda)v] \\ &= \begin{cases} 0 & \text{if } v \neq 0, \\ -\frac{1}{\lambda} \lim_{u \rightarrow 0^+} \lambda u \log(\lambda u) = 0 & \text{if } v = 0. \end{cases} \end{aligned}$$

Since $g_\lambda(p) = \sum_x s(p(x), \lambda p(x) + (1 - \lambda)q(x))$, the continuity of g_λ then follows. Since g_λ is also convex on $\mathcal{P}(\mathcal{X})$ by item (v) of Proposition 1, and the set $\mathcal{P}(\mathcal{X})$ is compact, the function g_λ takes its maximum at some extreme point δ_x of $\mathcal{P}(\mathcal{X})$. It follows that

$$\sup_{p \in \mathcal{P}(\mathcal{X})} S(p \| \lambda p + (1 - \lambda)q) = \max_{x \in \mathcal{X}} S(\delta_x \| \lambda \delta_x + (1 - \lambda)q) = \log \frac{1}{\lambda + (1 - \lambda) \min_{x \in \mathcal{X}} q(x)}.$$

Setting $q_{\min} = \min_{x \in \mathcal{X}} q(x)$, the derivative in λ of the last expression is

$$\frac{d}{d\lambda} \left(\log \frac{1}{\lambda + (1 - \lambda)q_{\min}} \right) = \frac{q_{\min} - 1}{(1 - q_{\min})\lambda + q_{\min}},$$

which is negative for all $\lambda \in (0, 1]$ since $q_{\min} \leq 1/|\mathcal{X}| < 1$. Thus, the right hand side of (3) is strictly decreasing in λ . \square

Note that, if $\lambda = 0$, then $g_0(p) = S(p \| q)$ is an extended real LSC function on $\mathcal{P}(\mathcal{X})$ by item (vi) of Proposition 1. However, it is not difficult to show along the lines of the previous proof that the maximum in (3) is still attained, and

$$\max_{p \in \mathcal{P}(\mathcal{X})} S(p \| q) = \begin{cases} \log \frac{1}{\min_x q(x)} & \text{if } \text{supp } q = \mathcal{X} \\ +\infty & \text{otherwise.} \end{cases}$$

2 Entropic measurement uncertainty relations

In general, the two reference observables A and B , introduced at the end of Section 1.1, are incompatible and only “approximate” joint measurements are possible for them [22, 23, 42, 43]. Moreover, any measurement of A may disturb a subsequent measurement of B , in a way that the resulting distribution of B can be very far from its unperturbed version; this disturbance may be present even when the two observables are compatible. Typically, such a disturbance of A on B can not be removed, nor just made arbitrarily small, unless we drop the requirement of exactly measuring A . However, in both cases, the measurement uncertainties on A and B can not always be made equally small. The quantum nature of A and B relates their measurement uncertainties, so that improving the approximation of A affects the quality of the corresponding approximation of B and vice versa. Incompatibility of A and B on the one hand, and the disturbance induced on B by a measurement of A on the other hand, are alternative manifestations of the quantum relation between the two observables [23], and as such deserve different approaches.

Our aim is now to quantify both these types of measurement uncertainty relations between A and B by means of suitable informational quantities. In the case of incompatible observables, we will produce an *entropic incompatibility degree*, encoding the minimum total error affecting any approximate joint measurement of A and B . Similarly, when the observable B is measured after an approximate version of A , the uncertainties resulting on both the observables will produce an *error/disturbance tradeoff* for A and B . In both cases, we will look for an optimal bi-observable M whose marginals $M_{[1]}$ and $M_{[2]}$ are the best approximations of the two target observables A and B . However, the different points of view will reflect on the fact that we will optimize over M in two different sets, according to the case at hand.

2.1 Entropic divergence for observables

We consider now any bi-observable $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ as an *approximate joint measurement* of A and B and we want an informational quantification of how far its marginals $M_{[1]}$ and $M_{[2]}$ are from correctly approximating the two reference observables A and B .

Following [5–7], these two approximations will be judged by comparing (within our entropic approach) the distribution $M_{[1]}^\rho$ with A^ρ , and the distribution $M_{[2]}^\rho$ with B^ρ , for different possible states ρ . Note that we can not compare the output of $M_{[1]}$ with that of A , and the output of $M_{[2]}$ with that of B , in one and the same experiment. Indeed, our bi-observable M is a joint measurement of $M_{[1]}$ and $M_{[2]}$, and there is no way to turn it into a joint measurement of the four observables A , $M_{[1]}$, $M_{[2]}$ and B , when A and B are not

compatible. Nevertheless, even if A and B are incompatible, each of them can be measured in independent repetitions of a preparation (state) ρ of the system. Similarly, any bi-observable M can be measured in other independent repetitions of the same preparation. So, all the three probability distributions A^ρ , B^ρ , M^ρ can be estimated from independent experiments and then they can be compared without any hypothesis of compatibility among A , B and M .

Our first step is to quantify the inefficiency of the distribution approximations $A^\rho \simeq M_{[1]}^\rho$ and $B^\rho \simeq M_{[2]}^\rho$, given the bi-observable M . According to the discussion in Section 1.2, the natural way to quantify the loss of information in such approximations is to use the relative entropy. Then, the total amount of lost information is the *error function*

$$S[A, B \| M](\rho) = S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho). \quad (4)$$

By Proposition 1, item (vii), we can rewrite (4) in the form

$$S[A, B \| M](\rho) = S(A^\rho \otimes B^\rho \| M_{[1]}^\rho \otimes M_{[2]}^\rho), \quad (5)$$

which, although lacking of any particular physical meaning (as the joint distribution $A^\rho \otimes B^\rho$ does not correspond to a single experiment on the system), however can be mathematically useful in some situations (see e.g. the proof of Theorem 8).

The second step is to quantify the inefficiency of the observable approximations $A \simeq M_{[1]}$ and $B \simeq M_{[2]}$ by means of the marginals of a given bi-observable M , without reference to any particular state. In order to construct a state-independent quantity, we take the worst case in (4) with respect to the system state ρ .

Definition 1. We call *entropic divergence* of $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ from (A, B) the quantity

$$D(A, B \| M) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} \left\{ S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \right\}. \quad (6)$$

Therefore, the entropic divergence $D(A, B \| M)$ quantifies the worst total loss of information due to the two approximations. It depends only on the marginals of M , and so it is the same for different bi-observables with equal marginals. Clearly, if A and B are compatible and M is any of their joint measurements, then $D(A, B \| M) = 0$ by Proposition 1, item (ii).

Note that there is a *unique* supremum over ρ in (6), so that the entropic divergence $D(A, B \| M)$ takes into account all possible compensations between the information lost in the first and in the second approximation. For this reason, our approach is not a mere entropic version of the point of view of Busch, Lahti, Werner, with the relative entropy replacing the Wasserstein (or transport) distance. Indeed, in [5–7, 11], for a given bi-observable M , the worst possible errors in the two approximations are evaluated separately and without balancing, and the two suprema for the Wasserstein distances $d(A^{\rho_1}, M_{[1]}^{\rho_1})$ and $d(B^{\rho_2}, M_{[2]}^{\rho_2})$ are attained at possibly different states ρ_1 and ρ_2 . However, from the resulting tradeoff between the maximal errors $d(A, M_{[1]}) = \sup_{\rho} d(A^\rho, M_{[1]}^\rho)$ and $d(B, M_{[2]}) = \sup_{\rho} d(B^\rho, M_{[2]}^\rho)$, one can not infer how well the the distribution M^ρ approximates the probabilities A^ρ and B^ρ when the observables are measured on the same state ρ . The difference of the two approaches reflects also in the fact that there are several cases in which the optimal approximate joint measurements saturating the bounds of [5–7, 11] are not unique, while there is only one bi-observable M minimizing the entropic divergence (6). Two orthogonal spin-1/2 observables and the whole family of MUB observables described in Sections 3.2 and 3.3 are examples of couples (A, B) sharing the latter property (see Proposition 14 and item 2 of Remark 8).

Theorem 3. Let $A \in \mathcal{M}(\mathcal{X})$, $B \in \mathcal{M}(\mathcal{Y})$ be the reference observables. For the entropic divergences defined above the following properties hold:

- (i) The function $S[A, B \| M] : \mathcal{S}(\mathcal{H}) \rightarrow [0, +\infty]$ is convex and LSC, $\forall M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$.
- (ii) The function $D(A, B \| \cdot) : \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \rightarrow [0, +\infty]$ is convex and LSC.
- (iii) For any $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, the following three statements are equivalent:
 - (a) $D(A, B \| M) < +\infty$,

(b) $\ker M_{[1]}(x) \subseteq \ker A(x) \forall x$ and $\ker M_{[2]}(y) \subseteq \ker B(y) \forall y$,

(c) $S[A, B||M]$ is bounded and continuous.

(iv) $D(A, B||M) = \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \left\{ S(A^\rho || M_{[1]}^\rho) + S(B^\rho || M_{[2]}^\rho) \right\}$, where the maximum can be any value in the interval $[0, +\infty]$.

(v) The entropic divergence $D(A, B||M)$ is invariant under an overall unitary conjugation of the observables A, B, M , and a relabelling of their outcome spaces \mathcal{X} and \mathcal{Y} .

Proof. (i) The function $S[A, B||M]$ is the sum of two terms which are convex, because the mapping $\rho \mapsto X^\rho$ is affine for any observable X and by item (v) of Proposition 1; hence $S[A, B||M]$ is convex. Moreover, each term is LSC, since $\rho \mapsto X^\rho$ is continuous and because of item (vi) of Proposition 1; so the sum $S[A, B||M]$ is LSC by [41, Prop. 1.5.12].

(ii) Each mapping $M \mapsto M_{[i]}^\rho$ is affine and continuous, and the functions $S(A^\rho || \cdot)$, $S(B^\rho || \cdot)$ are convex and LSC by Proposition 1, items (v) and (vi). It follows that $M \mapsto S(A^\rho || M_{[1]}^\rho)$ and $M \mapsto S(B^\rho || M_{[2]}^\rho)$ are also convex and LSC functions on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$; hence, such are their sum and the supremum $D(A, B||\cdot)$ [41, Prop. 1.5.12].

(iii) Let us show (a) \implies (b) \implies (c) \implies (a).

(a) \implies (b). If $\ker M_{[1]}(x) \not\subseteq \ker A(x)$ for some x , then we could take a pure state $\rho = |\psi\rangle\langle\psi|$ with ψ belonging to $\ker M_{[1]}(x)$ but not to $\ker A(x)$, so that $M_{[1]}^\rho(x) = 0$ while $A^\rho(x) > 0$, and thus we would get $S(A^\rho || M_{[1]}^\rho) = +\infty$ and the contradiction $D(A, B||M) = +\infty$.

(b) \implies (c). The function $S[A, B||M]$ is a finite sum of terms of the kind $s(A^\rho(x), M_{[1]}^\rho(x))$ or $s(B^\rho(y), M_{[2]}^\rho(y))$, where s is the function defined in (2). Under the hypothesis (b), each of these terms is a bounded and continuous function of ρ by Lemma 4 below. We thus conclude that $S[A, B||M]$ is bounded and continuous.

(c) \implies (a). Trivial, as $D(A, B||M) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} S[A, B||M](\rho)$.

(iv) If $D(A, B||M) < +\infty$, then $S[A, B||M]$ is a bounded and continuous function on the compact set $\mathcal{S}(\mathcal{H})$ by item (iii) above, and thus it attains a maximum; moreover, $S[A, B||M]$ is convex, hence it has at least a maximum point among the extreme points of $\mathcal{S}(\mathcal{H})$, which are the pure states. If instead $D(A, B||M) = +\infty$, then $\ker M_{[1]}(x) \not\subseteq \ker A(x)$ for some x , or $\ker M_{[2]}(y) \not\subseteq \ker B(y)$ for some y again by item (iii). In this case, every pure state $\rho = |\psi\rangle\langle\psi|$ with $\psi \in \ker M_{[1]}(x) \setminus \ker A(x)$, or $\psi \in \ker M_{[2]}(y) \setminus \ker B(y)$ is such that $S[A, B||M](\rho) = +\infty$, and thus is a maximum point of $S[A, B||M]$.

(v) The invariance under unitary conjugation follows from the fact that, for any unitary operator U on \mathcal{H} , we have $US(\mathcal{H})U^* = \mathcal{S}(\mathcal{H})$ and $(U^*MU)_{[i]} = U^*M_{[i]}U$, so that

$$\begin{aligned} & \sup_{\rho \in \mathcal{S}(\mathcal{H})} \left\{ S((U^*AU)^\rho || (U^*MU)_{[1]}^\rho) + S((U^*BU)^\rho || (U^*MU)_{[2]}^\rho) \right\} \\ &= \sup_{\rho \in \mathcal{S}(\mathcal{H})} \left\{ S(A^{U\rho U^*} || M_{[1]}^{U\rho U^*}) + S(B^{U\rho U^*} || M_{[2]}^{U\rho U^*}) \right\} \\ &= \sup_{\rho \in \mathcal{S}(\mathcal{H})} \left\{ S(A^\rho || M_{[1]}^\rho) + S(B^\rho || M_{[2]}^\rho) \right\}. \end{aligned}$$

The invariance under relabelling of the outcomes is an immediate consequence of the analogous property of the relative entropy (Proposition 1, item (iv)). \square

An essential step in the last proof is the following lemma.

Lemma 4. Suppose $A, B \in \mathcal{L}(\mathcal{H})$ are such that $0 \leq A \leq \mathbb{1}$ and $0 \leq B \leq \mathbb{1}$, and assume that $\ker B \subseteq \ker A$. Then, the function $s_{A,B} : \mathcal{S}(\mathcal{H}) \rightarrow [0, +\infty]$, with

$$s_{A,B}(\rho) = s(A^\rho, B^\rho) \quad \text{and} \quad A^\rho = \text{Tr}\{A\rho\}, \quad B^\rho = \text{Tr}\{B\rho\}$$

is bounded and continuous.

Proof. We will show that $s_{A,B}$ is a continuous function on $\mathcal{S}(\mathcal{H})$; since $\mathcal{S}(\mathcal{H})$ is compact, this will also imply that $s_{A,B}$ is bounded. The case $B = 0$ is trivial, hence we will suppose $B \neq 0$. By the hypotheses, the condition $B^\rho = 0$ implies that $A^\rho = 0$. The definition (2) of s then gives

$$s_{A,B}(\rho) = \begin{cases} A^\rho \log \frac{A^\rho}{B^\rho} & \text{if } A^\rho > 0 \text{ and } B^\rho > 0 \\ 0 & \text{if } A^\rho = 0 \text{ and } B^\rho > 0 \\ 0 & \text{if } A^\rho = 0 \text{ and } B^\rho = 0 \end{cases} = \begin{cases} B^\rho h\left(\frac{A^\rho}{B^\rho}\right) & \text{if } B^\rho > 0 \\ 0 & \text{if } B^\rho = 0 \end{cases}$$

where we introduced the continuous function $h : [0, +\infty) \rightarrow [-1/2, +\infty)$, with

$$h(t) = \begin{cases} t \log t & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{cases}$$

The function $s_{A,B}$ is clearly continuous on the open subset $\mathcal{U} = \{\rho \in \mathcal{S}(\mathcal{H}) : B^\rho > 0\}$ of the state space $\mathcal{S}(\mathcal{H})$. It remains to show that it is also continuous at all the points of the set $\mathcal{U}^c = \{\rho \in \mathcal{S}(\mathcal{H}) : B^\rho = 0\}$. To this aim, observe that

$$A \leq c_{\max}(A)P_A \leq c_{\max}(A)P_B \quad \text{and} \quad B \geq c_{\min}(B)P_B,$$

where $c_{\max}(A)$ is the maximum eigenvalue of A , $c_{\min}(B)$ is the minimum positive eigenvalue of B , and we denote by P_A and P_B the orthogonal projections onto $\ker A^\perp$ and $\ker B^\perp$, respectively. Since $P_B^\rho \neq 0$ for all ρ such that $B^\rho > 0$, it follows that

$$0 \leq \frac{A^\rho}{B^\rho} \leq \frac{c_{\max}(A)}{c_{\min}(B)}, \quad \forall \rho \in \mathcal{U},$$

hence, by continuity of h and boundedness of the interval $[0, c_{\max}(A)/c_{\min}(B)]$, there is a constant $M > 0$ such that

$$|s_{A,B}(\rho)| = \left| B^\rho h\left(\frac{A^\rho}{B^\rho}\right) \right| \leq MB^\rho, \quad \forall \rho \in \mathcal{U}.$$

On the other hand, for $\rho \in \mathcal{U}^c$ we have $s_{A,B}(\rho) = 0$. If $(\rho_k)_k$ is a sequence in $\mathcal{S}(\mathcal{H})$ converging to $\rho_0 \in \mathcal{U}^c$, then $|s_{A,B}(\rho_k) - s_{A,B}(\rho_0)| \leq MB^{\rho_k} \xrightarrow[k \rightarrow \infty]{} 0$, which shows that $s_{A,B}$ is continuous at ρ_0 . \square

2.2 Incompatibility degree, error/disturbance tradeoff and optimal approximate joint measurements

After introducing the error function $S[A, B \| M](\rho)$, which describes the total information lost by measuring the bi-observable M in place of A and B in the state ρ , and after defining its maximum value $D(A, B \| M)$ over all states, now the third step is to quantify the intrinsic measurement uncertainties between A and B , dropping any reference to a particular state or approximating joint measurement. When we are interested in incompatibility, this is done by taking the minimum of the divergence $D(A, B \| M)$ over all possible bi-observables $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$. The resulting quantity is the minimum inefficiency which can not be avoided when the (possibly incompatible) observables A and B are approximated by the compatible marginals $M_{[1]}$ and $M_{[2]}$ of any bi-observable M . This minimum can be understood as an “incompatibility degree” of the two observables A and B .

Definition 2. The *entropic incompatibility degree* $c_{\text{inc}}(A, B)$ of the observables A and B is

$$c_{\text{inc}}(A, B) = \inf_{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} D(A, B \| M). \quad (7)$$

The definition is consistent, as obviously $c_{\text{inc}}(A, B) \geq 0$, and $c_{\text{inc}}(A, B) = 0$ when A and B are compatible.

As the notion of incompatibility is symmetric by exchanging the observables A and B , we would expect that also the incompatibility degree satisfies the property $c_{\text{inc}}(A, B) = c_{\text{inc}}(B, A)$. Indeed, this is actually

true, as $D(A, B \| M) = D(A, B \| M')$ for all $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, where $M' \in \mathcal{M}(\mathcal{Y} \times \mathcal{X})$ is defined by $M'(y, x) = M(x, y)$. Note that the symmetry of c_{inc} comes from the fact that, in defining the error function $S[A, B \| M]$, we chose equal weights for the contributions of the two approximation errors of A and B.

On the other hand, when we deal with the error/disturbance uncertainty relation, our analysis is restricted to the bi-observables describing sequential measurements of an approximate version A' of A, followed by an exact measurement of B. In other words, we focus on the subset of $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$

$$\begin{aligned} \mathcal{M}(\mathcal{X}; B) &= \{\mathcal{J}^*(B) : \mathcal{J} \in \mathcal{J}(\mathcal{X})\} \\ &= \{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) : M(x, y) = \mathcal{J}_x^*[B(y)] \ \forall x, y, \text{ for some } \mathcal{J} \in \mathcal{J}(\mathcal{X})\}, \end{aligned} \quad (8)$$

consisting of the sequential measurements where the first outcome set \mathcal{X} and the second observable B are fixed. If $M = \mathcal{J}^*(B) \in \mathcal{M}(\mathcal{X}; B)$, then $A' = M_{[1]} = \mathcal{J}^*[\mathbb{1}]$ is the observable approximating A, and $B' = \mathcal{J}_x^*[B(\cdot)]$ is the version of B perturbed as an effect of the measurement of A' . In general, it may equally well be $A' \neq A$ and $B' \neq B$, unless the observable A can be measured without disturbing B [23].

In order to quantify the measurement uncertainties due to the error/disturbance tradeoff, we then consider the minimum of the entropic divergence $D(A, B \| M)$ for $M \in \mathcal{M}(\mathcal{X}; B)$. If we read $S(A^\rho \| M_{[1]}^\rho)$ as the error made by \mathcal{J} in measuring A in the state ρ , and $S(B^\rho \| M_{[2]}^\rho)$ as the amount of disturbance introduced by \mathcal{J} on the subsequent measurement of B, then the divergence $D(A, B \| M)$ expresses the sum error + disturbance maximized over all states for the sequential measurement M. Minimizing $D(A, B \| M)$ over all sequential measurements, we then obtain the following entropic quantification of the error/disturbance uncertainty relation between A and B.

Definition 3. The entropic error/disturbance coefficient $c_{\text{ed}}(A, B)$ of A followed by B is

$$c_{\text{ed}}(A, B) = \inf_{M \in \mathcal{M}(\mathcal{X}; B)} D(A, B \| M). \quad (9)$$

Similarly to the incompatibility degree, the error/disturbance coefficient is always nonnegative, and $c_{\text{ed}}(A, B) = 0$ when A can be measured without disturbing B. Contrary to c_{inc} , we stress that in general the two indexes $c_{\text{ed}}(A, B)$ and $c_{\text{ed}}(B, A)$ can be different, as shown in Remark 2 below.

When the approximate measurement of the first observable A is described by the instrument \mathcal{J} , the measurement of the second fixed observable B could be preceded by any kind of correction taking into account the observed outcome x [6]. This can be formalized by inserting a quantum channel \mathcal{C}_x in between the measurements of A and B. As the composition $\mathcal{J}'_x = \mathcal{C}_x \circ \mathcal{J}_x$ gives again an instrument $\mathcal{J}' \in \mathcal{J}(\mathcal{X})$, we then see that any possible correction is considered when we take the infimum in $\mathcal{M}(\mathcal{X}; B)$. The latter fact shows that Definition 3 is consistent, since only by taking into account all possible corrections we can properly speak of pure unavoidable disturbance and of error/disturbance tradeoff.

Comparing the two indexes c_{inc} and c_{ed} , the inequality $c_{\text{inc}}(A, B) \leq c_{\text{ed}}(A, B)$ trivially follows from the inclusion $\mathcal{M}(\mathcal{X}; B) \subseteq \mathcal{M}(\mathcal{X} \times \mathcal{Y})$. This means that, even if one is interested in c_{ed} , the most symmetric index c_{inc} is at least a lower bound for the other coefficient.

We stress that the inclusion $\mathcal{M}(\mathcal{X}; B) \subseteq \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ may be strict in general. For example, there may exist observables which are compatible with B, but can not be measured before B without disturbing it. Then, taken such an observable A, a joint measurement of A and B clearly belongs to $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ but can not be in $\mathcal{M}(\mathcal{X}; B)$. When $\mathcal{M}(\mathcal{X}; B) \subsetneq \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, the incompatibility and error/disturbance approaches definitely are not equivalent. Nevertheless, there is one remarkable situation in which the two approaches are the same.

Proposition 5. If $B \in \mathcal{M}(\mathcal{Y})$ is a sharp observable, then $\mathcal{M}(\mathcal{X}; B) = \mathcal{M}(\mathcal{X} \times \mathcal{Y})$.

Proof. The proof directly follows from the argument at the end of [23, Sect. II.D]. Indeed, for any $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we can define the instrument $\mathcal{J} \in \mathcal{J}(\mathcal{X})$ with

$$\mathcal{J}_x[\rho] = \sum_{y \in \mathcal{Y} : B(y) \neq 0} \text{Tr} \{ \rho M(x, y) \} \frac{B(y)}{\text{Tr} \{ B(y) \}}.$$

For such an instrument, the equality $M(x, y) = \mathcal{J}_x^*[B(y)]$ is immediate. \square

As an immediate consequence of this result, we have $c_{\text{ed}}(A, B) = c_{\text{inc}}(A, B)$ whenever the second measured observable B is sharp.

By Theorem 6 below, the two coefficients $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ are finite, and the two infima in their definitions are actually two minima. It is convenient to give a name to the corresponding sets of minimizing bi-observables:

$$\mathcal{M}_{\text{inc}}(A, B) = \arg \min_{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} D(A, B \| M), \quad \mathcal{M}_{\text{ed}}(A, B) = \arg \min_{M \in \mathcal{M}(\mathcal{X}; B)} D(A, B \| M). \quad (10)$$

We can say that $\mathcal{M}_{\text{inc}}(A, B)$ is the set of the *optimal approximate joint measurements* of A and B . Similarly, $\mathcal{M}_{\text{ed}}(A, B)$ contains the sequential measurements optimally approximating A and B .

The next theorem summarizes the main properties of c_{inc} and c_{ed} contained in the above discussion, and states some further relevant facts about the two indexes.

Theorem 6. *Let $A \in \mathcal{M}(\mathcal{X})$, $B \in \mathcal{M}(\mathcal{Y})$ be the reference observables. For the entropic coefficients defined above the following properties hold:*

- (i) *The coefficients $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ are invariant under an overall unitary conjugation of the observables A and B , and do not depend on the labelling of the outcomes in \mathcal{X} and \mathcal{Y} .*
- (ii) *The incompatibility degree has the exchange symmetry property $c_{\text{inc}}(A, B) = c_{\text{inc}}(B, A)$.*
- (iii) *$c_{\text{inc}}(A, B) \leq c_{\text{ed}}(A, B) \leq \log |\mathcal{X}| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} H(A^\rho)$ and $c_{\text{inc}}(A, B) \leq \log |\mathcal{Y}| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} H(B^\rho)$.*
- (iv) *The sets $\mathcal{M}_{\text{inc}}(A, B)$ and $\mathcal{M}_{\text{ed}}(A, B)$ are nonempty convex compact subsets of $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$.*
- (v) *$c_{\text{inc}}(A, B) = 0$ if and only if the observables A and B are compatible (jointly measurable), and in this case $\mathcal{M}_{\text{inc}}(A, B)$ is the set of all the joint measurements of A and B .*
- (vi) *$c_{\text{ed}}(A, B) = 0$ if and only if the observables A and B are sequentially compatible, and in this case $\mathcal{M}_{\text{ed}}(A, B)$ is the set of all the sequential joint measurements of A and B .*
- (vii) *If B is a sharp observable, then $c_{\text{inc}}(A, B) = c_{\text{ed}}(A, B)$ and $\mathcal{M}_{\text{inc}}(A, B) = \mathcal{M}_{\text{ed}}(A, B)$.*

Proof. (i) The invariance under unitary conjugation follows from the corresponding property of the entropic divergence (Theorem 3, item (v)). We will prove it only for c_{ed} , the case of c_{inc} being even simpler. We have

$$c_{\text{ed}}(U^*AU, U^*BU) = \inf_{M \in \mathcal{M}(\mathcal{X}; U^*BU)} D(U^*AU, U^*BU \| M) = \inf_{M' \in U\mathcal{M}(\mathcal{X}; U^*BU)U^*} D(A, B \| M'),$$

and, in order to show that $c_{\text{ed}}(U^*AU, U^*BU) = c_{\text{ed}}(A, B)$, it only remains to prove the set equality $U\mathcal{M}(\mathcal{X}; U^*BU)U^* = \mathcal{M}(\mathcal{X}; B)$. If $M = \mathcal{J}^*(U^*BU) \in \mathcal{M}(\mathcal{X}; U^*BU)$, then, defining the instrument $\mathcal{J}'_x[\rho] = U\mathcal{J}_x[U^*\rho U]U^*$, $\forall \rho, x$, we have $UMU^* = \mathcal{J}'^*(B) \in \mathcal{M}(\mathcal{X}; B)$, as claimed. In a similar way, the invariance under relabelling of the outcomes is a consequence of the analogous property of the entropic divergence.

(ii) This property has already been noticed.

(iii) The first inequality follows from the inclusion $\mathcal{M}(\mathcal{X}; B) \subseteq \mathcal{M}(\mathcal{X} \times \mathcal{Y})$. Then, let $\mathcal{U} \in \mathcal{J}(\mathcal{X})$ be the trivial uniform instrument $\mathcal{U}_x[\rho] = u_{\mathcal{X}}(x)\rho$, $u_{\mathcal{X}}(x) = 1/|\mathcal{X}|$. Taking the sequential measurement $\mathcal{U}^*(B) \in \mathcal{M}(\mathcal{X}; B)$, we get $\mathcal{U}^*(B)^\rho = u_{\mathcal{X}} \otimes B^\rho$ and

$$S(A^\rho \| \mathcal{U}^*(B)^\rho_{[1]}) + S(B^\rho \| \mathcal{U}^*(B)^\rho_{[2]}) = S(A^\rho \| u_{\mathcal{X}}) = \log |\mathcal{X}| - H(A^\rho).$$

By taking the supremum over all the states, we get $D(A, B \| \mathcal{U}^*(B)) = \log |\mathcal{X}| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} H(A^\rho)$, hence $c_{\text{ed}}(A, B) \leq \log |\mathcal{X}| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} H(A^\rho)$ by definition. The last inequality then follows by the exchange symmetry property of $c_{\text{inc}}(A, B)$.

(iv) By item (ii) in Theorem 3 and item (iii) just above, $D(A, B \| \cdot)$ is a convex LSC proper (i.e. not identically $+\infty$) function on the compact set $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$. This implies that $\mathcal{M}_{\text{inc}}(A, B) \neq \emptyset$ [41, Exercise

E.1.6]. Closedness and convexity of $\mathcal{M}_{\text{inc}}(A, B)$ are then easy and standard consequences of $D(A, B \| \cdot)$ being convex and LSC. On the other hand, the set $\mathcal{M}(\mathcal{X}; B)$ is a convex and compact subset of $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$; indeed, this follows from convexity and compactness of $\mathcal{J}(\mathcal{X})$ and continuity of the mapping $\mathcal{J} \mapsto \mathcal{J}^*(B)$ in the definition (8). The proof that the subset $\mathcal{M}_{\text{ed}}(A, B) \subseteq \mathcal{M}(\mathcal{X}; B)$ is nonempty, convex and compact then follows along the same lines of $\mathcal{M}_{\text{inc}}(A, B)$.

(v) Assume $c_{\text{inc}}(A, B) = 0$. Then $\mathcal{M}_{\text{inc}}(A, B)$ exactly consists of all the joint measurements of A and B , which therefore turn out to be compatible, as $\mathcal{M}_{\text{inc}}(A, B) \neq \emptyset$ by (iv). Indeed, if $M \in \mathcal{M}_{\text{inc}}(A, B)$, then $0 = c_{\text{inc}}(A, B) = D(A, B \| M)$, which gives $S(A^\rho \| M_{[1]}^\rho) = S(B^\rho \| M_{[2]}^\rho) = 0$ for all ρ . By item (ii) of Proposition 1, this yields $A^\rho = M_{[1]}^\rho$, $B^\rho = M_{[2]}^\rho$, $\forall \rho$, and so $A = M_{[1]}$, $B = M_{[2]}$, which means that M is a joint measurement of A and B . The converse implication was already noticed in the text.

(vi) Similarly to the previous item, if $c_{\text{ed}}(A, B) = 0$, then $\mathcal{M}_{\text{ed}}(A, B)$ consists exactly of all the sequential joint measurements of A and B . Indeed, by the same argument of (v), if $M \in \mathcal{M}_{\text{ed}}(A, B)$, then M is a joint measurement of A and B ; but now $M \in \mathcal{M}_{\text{ed}}(A, B) \subseteq \mathcal{M}(\mathcal{X}; B)$, which means that M is also a sequential measurement. Since $\mathcal{M}_{\text{ed}}(A, B) \neq \emptyset$ by (iv), the condition $c_{\text{ed}}(A, B) = 0$ implies that A and B are sequentially compatible. The other implication is trivial and was already remarked.

(vii) As observed above, if B is sharp, then by Proposition 5 we have $\mathcal{M}(\mathcal{X}; B) = \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, which implies the claim. \square

We see in items (v) and (vi) that the two indexes c_{inc} and c_{ed} have the desirable feature of being zero exactly when the two observables A and B satisfy the corresponding compatibility or nondisturbance property. We also stress that, by their very definitions, $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ are independent of both the prepared states ρ and the approximating bi-observables M , as well as they satisfy the natural invariance properties of item (i) of Theorem 6. In view of these facts, we are allowed once more to say that $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ are proper quantifications of the intrinsic incompatibility and error/disturbance affecting the two observables A and B .

Remark 1 (MUR). By definition, the two coefficients (7) and (9) are lower bounds for the entropic divergence (6) of every bi-observable M from (A, B) :

$$D(A, B \| M) \geq c_{\text{inc}}(A, B), \quad \forall M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}); \quad D(A, B \| M) \geq c_{\text{ed}}(A, B), \quad \forall M \in \mathcal{M}(\mathcal{X}; B). \quad (11)$$

These inequalities are two entropic MUR for the reference observables A and B . By properties (v) and (vi) of Theorem 6, the two MUR are non trivial and, by property (iv), both the bounds are tight. By property (iv) of Theorem 3, we have also the following formulation of the uncertainty relations:

$$\forall M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}), \quad \exists \rho \in \mathcal{S}(\mathcal{H}) : S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \geq c_{\text{inc}}(A, B); \quad (12)$$

$$\forall M \in \mathcal{M}(\mathcal{X}; B), \quad \exists \rho \in \mathcal{S}(\mathcal{H}) : S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \geq c_{\text{ed}}(A, B). \quad (13)$$

So, in an approximate joint measurement of A and B , the total loss of information can not be arbitrarily reduced: it depends on the state ρ but, for sure, it can be as big as $c_{\text{inc}}(A, B)$. Similarly, in a sequential measurement of A and B , there is a tradeoff between the information lost in the first measurement (because of the approximation error) and the information lost in the second measurement (because of the disturbance): they both depend on the state ρ but, for sure, their sum can be as big as $c_{\text{ed}}(A, B)$.

However, let us note that the definitions of the two bounds $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$ are rather implicit. Indeed, even if we proved that they are strictly positive when A and B are incompatible (or sequentially incompatible), their evaluation requires the two optimizations “sup” on the states and “inf” on the measurements. Nevertheless, in some cases explicit computations are possible (even including the evaluation of the optimal approximate joint measurements) or explicit lower bounds can be exhibited, see Sections 3.2 and 3.3.

Remark 2. Property (vii) of Theorem 6 says that the two indexes coincide in the important case in which B is sharp. However, this is not true in general, as shown e.g. by the two examples in Appendix A (taken from [23]). In the first example, $\dim \mathcal{H} = 3$, $|\mathcal{X}| = 2$, $|\mathcal{Y}| = 5$, and we have $c_{\text{ed}}(A, B) > c_{\text{ed}}(B, A) = c_{\text{inc}}(A, B) = 0$. The second example is more symmetric and simpler ($|\mathcal{X}| = |\mathcal{Y}| = 2$), and it yields $c_{\text{ed}}(A, B) > c_{\text{inc}}(A, B) = 0$ and also $c_{\text{ed}}(B, A) > 0$.

2.3 Noisy observables and uncertainty upper bounds

Before attempting an exact computation or at least trying to determine a lower bound for our coefficients $c_{\text{inc}}(A, B)$ and $c_{\text{ed}}(A, B)$, let us look for a better upper bound than the one provided in Theorem 6, item (iii). For this task, we introduce an important class of bi-observables M that are known to give good approximations of A and B . Even if these M were not optimal in our entropic approach, we expect that they should have a small divergence from (A, B) and thus they should give a good upper bound for its minimum.

Two incompatible observables A and B can always be turned into a compatible pair by adding enough classical noise to their measurements. Indeed, for any choice of trivial observables $T_A = p_A \mathbb{1}$, $p_A \in \mathcal{P}(\mathcal{X})$, and $T_B = p_B \mathbb{1}$, $p_B \in \mathcal{P}(\mathcal{Y})$, the observables $\lambda A + (1 - \lambda)T_A$ and $\gamma B + (1 - \gamma)T_B$, which are *noisy versions* of A and B with *noise intensities* $1 - \lambda$ and $1 - \gamma$, are compatible for all $\lambda, \gamma \in [0, 1]$ such that $\lambda + \gamma \leq 1$ (sufficient condition) [44, Prop. 1]. A joint measurement with the given marginals is

$$M(x, y) = \lambda A(x)p_B(y) + \gamma p_A(x)B(y) + (1 - \lambda - \gamma)p_A(x)p_B(y)\mathbb{1}.$$

Thus, any joint measurement M as above with $\lambda = \gamma = 1/2$ gives an upper bound for the incompatibility coefficient $c_{\text{inc}}(A, B)$:

$$\begin{aligned} c_{\text{inc}}(A, B) &\leq D(A, B \| M) = \sup_{\rho} \left\{ S\left(A^{\rho} \left\| \frac{1}{2} A^{\rho} + \frac{1}{2} p_A\right.\right) + S\left(B^{\rho} \left\| \frac{1}{2} B^{\rho} + \frac{1}{2} p_B\right.\right) \right\} \\ &\leq \sup_{\rho} S\left(A^{\rho} \left\| \frac{1}{2} A^{\rho} + \frac{1}{2} p_A\right.\right) + \sup_{\rho} S\left(B^{\rho} \left\| \frac{1}{2} B^{\rho} + \frac{1}{2} p_B\right.\right) \\ &\leq \log \frac{2}{1 + \min_x p_A(x)} + \log \frac{2}{1 + \min_y p_B(y)} \quad \text{by (3)} \\ &\leq 2. \end{aligned}$$

Note that the bound $c_{\text{inc}}(A, B) \leq 2$ does not depend neither on the observables A and B , nor on the Hilbert space dimension d .

Better upper bounds for $c_{\text{inc}}(A, B)$ can be evaluated along the above lines by considering specific p_A, p_B and going outside the region $\lambda + \gamma \leq 1$, and so reducing the noise intensities. For this purpose, for every $0 \leq \lambda \leq 1$, let us consider the couple of equally noisy observables

$$\begin{aligned} A_{\lambda}(x) &= \lambda A(x) + (1 - \lambda)A^{\rho_0}(x)\mathbb{1}, \\ B_{\lambda}(y) &= \lambda B(y) + (1 - \lambda)B^{\rho_0}(y)\mathbb{1}, \end{aligned} \tag{14}$$

where $\rho_0 = \frac{1}{d}\mathbb{1}$ is the maximally chaotic state. We have seen that, if $\lambda \leq 1/2$, then the two observables are always compatible, but, depending on the specific A and B , they could be compatible also for larger λ . In this case, the previous computation yields the upper bounds

$$c_{\text{inc}}(A, B) \leq D(A, B \| M) \leq \log \frac{1}{\lambda + (1 - \lambda) \min_{x \in \mathcal{X}} A^{\rho_0}(x)} + \log \frac{1}{\lambda + (1 - \lambda) \min_{y \in \mathcal{Y}} B^{\rho_0}(y)} \tag{15}$$

for all $\lambda \in [0, 1]$ such that A_{λ} and B_{λ} are compatible, and any joint measurement M of A_{λ} and B_{λ} . Since the two terms in the right hand side of (15) are decreasing functions of λ , in order to obtain the best bound we are thus led to find the maximal value λ_{max} of λ for which the noisy observables A_{λ} and B_{λ} are compatible. This problem was addressed in [45], where a complete solution was given for a couple of Fourier conjugate sharp observables. Moreover, it was shown that in the general case a nontrivial lower bound for λ_{max} can always be achieved by means of *optimal approximate cloning* [46].

Following the same idea, we are going to find a nontrivial upper bound for $c_{\text{inc}}(A, B)$ by means of the optimal approximate 2-cloning channel

$$\Phi : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H} \otimes \mathcal{H}), \quad \Phi(\rho) = \frac{2}{d+1} S_2(\rho \otimes \mathbb{1}) S_2,$$

where $S_2 : \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{H} \otimes \mathcal{H}$ is the orthogonal projection of $\mathcal{H} \otimes \mathcal{H}$ onto its symmetric subspace $\text{Sym}(\mathcal{H} \otimes \mathcal{H})$, defined by $S_2(\phi_1 \otimes \phi_2) = (\phi_1 \otimes \phi_2 + \phi_2 \otimes \phi_1)/2$. Performing a measurement of the product observable $A \otimes B$ in the state $\Phi(\rho)$ then amounts to measure the bi-observable $M_{\text{cl}} = \Phi^*(A \otimes B)$ in ρ ; its marginals are (see [47])

$$M_{\text{cl}}[1] = A_{\lambda_{\text{cl}}} \quad \text{and} \quad M_{\text{cl}}[2] = B_{\lambda_{\text{cl}}} \quad \text{where} \quad \lambda_{\text{cl}} = \frac{d+2}{2(d+1)}.$$

Of course $\lambda_{\text{cl}} \leq \lambda_{\text{max}}$, but the important point is that $\lambda_{\text{cl}} > 1/2$. Inserting the above λ_{cl} in the bound (15), we thus obtain

$$c_{\text{inc}}(A, B) \leq D(A, B \| M_{\text{cl}}) \leq \log \frac{2(d+1)}{d+2+d \min_x A^{\rho_0}(x)} + \log \frac{2(d+1)}{d+2+d \min_y B^{\rho_0}(y)},$$

holding for all observables A and B .

It is worth noticing that the bi-observable M_{cl} describes a sequential measurement having B as second measured observable. Indeed, define the instrument $\mathcal{J} \in \mathcal{J}(\mathcal{X})$, with

$$\mathcal{J}_x[\rho] = \text{Tr}_1 \{ (A(x) \otimes \mathbb{1}) \Phi(\rho) \},$$

where Tr_1 denotes the partial trace with respect to the first factor. It is easy to check that $M_{\text{cl}} = \mathcal{J}^*(B)$, so that $M_{\text{cl}} \in \mathcal{M}_{\text{ed}}(\mathcal{X}; B)$. Therefore, the upper bound we have found for $D(A, B \| M_{\text{cl}})$ actually provides a bound also for the entropic error/disturbance coefficient $c_{\text{ed}}(A, B)$.

Summarizing the above discussion, we thus arrive at the main conclusion of this section.

Theorem 7. *For any couple of observables A and B , we have*

$$c_{\text{inc}}(A, B) \leq c_{\text{ed}}(A, B) \leq \log \frac{2(d+1)}{d+2+d \min_x A^{\rho_0}(x)} + \log \frac{2(d+1)}{d+2+d \min_y B^{\rho_0}(y)} \leq 2. \quad (16)$$

The striking result is that the two uncertainty indexes lie between 0 and 2, independently of the Hilbert space dimension d and the number of possible outcomes $|\mathcal{X}|$ and $|\mathcal{Y}|$.

Remark 3. In the particular case in which both A and B are sharp, with $|\mathcal{X}| = |\mathcal{Y}| = d$ and $\text{rank } A(x) = \text{rank } B(y) = 1$ for all x, y , the probabilities A^{ρ_0} and B^{ρ_0} are the uniform measures on the respective outcome spaces: $A^{\rho_0} = u_{\mathcal{X}}$, $B^{\rho_0} = u_{\mathcal{Y}}$. The above inequalities then become

$$c_{\text{inc}}(A, B) = c_{\text{ed}}(A, B) \leq 2 \log \frac{2(d+1)}{d+3}, \quad (17)$$

which tends to 2 from below as $d \rightarrow \infty$.

For sharp observables the bound (16) is much better than the bound given in item (iii) of Theorem 6. However, the case of two trivial uniform observables $A = u_{\mathcal{X}} \mathbb{1}$ and $B = u_{\mathcal{Y}} \mathbb{1}$ is an example where the bound of Theorem 6 is better than the bound (16).

As a final consideration, we will later show that there are observables A and B such that their compatible noisy versions (14) do not optimally approximate A and B . Equivalently, for these observables all the elements $M \in \mathcal{M}_{\text{inc}}(A, B)$ (or $M \in \mathcal{M}_{\text{ed}}(A, B)$) have marginals $M_{[1]} \neq A_{\lambda}$ and $M_{[2]} \neq B_{\lambda}$ for all $\lambda \in [0, 1]$. Indeed, an example is provided by the two nonorthogonal sharp spin-1/2 observables in Section 3.2. The motivation of this feature comes from the fact that we are not making any extra assumption about our approximate joint measurements, as we optimize over the whole sets $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ or $\mathcal{M}(\mathcal{X}; B)$, according to the case at hand. This is the main difference with the approach, e.g., of [43, 45], where a degree of compatibility is defined by considering the minimal noise which one needs to add to A and B in order to make them compatible. It should also be remarked that the non-optimality of the noisy versions is true also in other contexts [42].

2.4 Connections with the preparation uncertainty

The entropic incompatibility degree and error/disturbance coefficient are the non trivial and tight lower bounds of the entropic MUR proved in Section 2.2. As we recalled in the Introduction, MUR are different from PUR, which have been formulated in the informational-theoretic framework by using different types of entropies (Shannon, Rényi, ...) [14–18]. Here we consider only the Shannon entropy and, to facilitate the connections with our indexes, we introduce the entropic *preparation uncertainty coefficient*

$$c_{\text{prep}}(A, B) = \inf_{\rho \in \mathcal{S}(\mathcal{H})} [H(A^\rho) + H(B^\rho)]. \quad (18)$$

According to the previous sections, the reference observables A and B are general POVMs. With this definition, the lower bound proved in [16, Cor. 2.6] can be written as

$$c_{\text{prep}}(A, B) \geq -\log \max_{x \in \mathcal{X}, y \in \mathcal{Y}} \|A(x)^{1/2} B(y)^{1/2}\|^2. \quad (19)$$

When the observables are sharp, this lower bound reduces to the one conjectured in [14] and proved in [15].

Note that the infimum in (18) actually is a minimum, because the two entropies are continuous in ρ . Moreover, the equality $c_{\text{prep}}(A, B) = 0$ is attained if and only if there exist two outcomes x and y such that both the positive operators $A(x)$ and $B(y)$ have at least *one* common eigenvector with eigenvalue 1.

For sharp observables, we immediately deduce that the absence of measurement uncertainty implies the absence of preparation uncertainty. Indeed, $c_{\text{inc}}(A, B) = 0$ is the same as A and B being compatible, which in turn is equivalent to the existence of a whole basis of common eigenvectors $\{\psi_i : i = 1, \dots, d\}$ for which both the distributions $\langle \psi_i | A(x) \psi_i \rangle$ and $\langle \psi_i | B(y) \psi_i \rangle$ reduce to Kronecker deltas [48, Cor. 5.3]. Therefore, we have the implication $c_{\text{inc}}(A, B) = 0 \implies c_{\text{prep}}(A, B) = 0$. However, the same relation fails for generic POVMs: for any couple of trivial observables A and B such that $A \neq \delta_x \mathbb{1}$ or $B \neq \delta_y \mathbb{1}$, we have $c_{\text{inc}}(A, B) = 0$ and $c_{\text{prep}}(A, B) > 0$. On the converse direction, the example of two non commuting sharp observables with a common eigenspace shows that in general $c_{\text{prep}}(A, B) = 0 \not\Rightarrow c_{\text{inc}}(A, B) = 0$. The failure of this implication exhibits a striking difference between the preparation and the measurement uncertainties: actually, the entropic incompatibility degree vanishes if and only if the two observables are compatible (Theorem 6, item (v)), while in the preparation case nothing similar happens.

Nevertheless, there exists a link between the entropic incompatibility degree c_{inc} and the preparation uncertainty coefficient c_{prep} . Indeed, let us consider the trivial uniform observable $U \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, with

$$U = (u_{\mathcal{X}} \otimes u_{\mathcal{Y}}) \mathbb{1} \quad \text{and} \quad U_{[1]} = u_{\mathcal{X}} \mathbb{1}, \quad U_{[2]} = u_{\mathcal{Y}} \mathbb{1}.$$

By property (iii) of Proposition 1, we have

$$S(A^\rho \| U_{[1]}^\rho) + S(B^\rho \| U_{[2]}^\rho) = \log |\mathcal{X}| + \log |\mathcal{Y}| - H(A^\rho) - H(B^\rho).$$

By taking the supremum over all the states, Definitions 1 and 2 give

$$c_{\text{inc}}(A, B) \leq D(A, B \| U) = \log |\mathcal{X}| + \log |\mathcal{Y}| - c_{\text{prep}}(A, B)$$

and the final result is

$$c_{\text{inc}}(A, B) + c_{\text{prep}}(A, B) \leq \log |\mathcal{X}| + \log |\mathcal{Y}|. \quad (20)$$

Note that this bound is saturated at least in the trivial case $A = u_{\mathcal{X}} \mathbb{1}$, $B = u_{\mathcal{Y}} \mathbb{1}$, for which we have $c_{\text{prep}}(A, B) = \log |\mathcal{X}| + \log |\mathcal{Y}|$ and $c_{\text{inc}}(A, B) = 0$. We also remark that (20) is not the trivial sum of the two upper bounds $c_{\text{inc}}(A, B) \leq 2$ (Theorem 7) and $c_{\text{prep}}(A, B) \leq \log |\mathcal{X}| + \log |\mathcal{Y}|$ (following from the definition (18) of c_{prep} and the bound for the Shannon entropy of Proposition 1, item (i)).

3 Symmetries and uncertainty lower bounds

Many fundamental observables in quantum mechanics are directly related to symmetry properties of the quantum system at hand. That is, in many concrete situations there is some symmetry group G acting on

both the measurement outcome space and the set of the quantum system states, in such a way that the two group actions naturally intertwine. The observables that preserve the symmetry structure are usually called *G-covariant*.

In the present setting, covariance will help us to prove a lower bound for the incompatibility degree $c_{\text{inc}}(A, B)$ and to characterize the optimal set $\mathcal{M}_{\text{inc}}(A, B)$ for a couple of sharp observables A and B sharing suitable symmetry properties. In Section 3.1 below we provide a general result in this sense, which we then apply to the cases of two spin-1/2 components (Section 3.2) and two observables that are conjugated by the Fourier transform of a finite field (Section 3.3).

3.1 Symmetries and optimal approximate joint measurements

We now suppose that the joint outcome space $\mathcal{X} \times \mathcal{Y}$ carries the action of a finite group G , acting on the left, so that each $g \in G$ is associated with a bijective map on the finite set $\mathcal{X} \times \mathcal{Y}$. Moreover, we also assume that there is a projective unitary representation U of G on \mathcal{H} . The following natural left actions are then defined for all $g \in G$:

- on $\mathcal{S}(\mathcal{H})$: $g\rho = U(g)\rho U(g)^*$;
- on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$: $gp(x, y) = p(g^{-1}(x, y))$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$;
- on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$: $gM(x, y) = U(g)M(g^{-1}(x, y))U(g)^*$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

While the two actions on $\mathcal{S}(\mathcal{H})$ and $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ have a clear physical interpretation, the action on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is understood by means of the fundamental relation

$$g(M^\rho) = (gM)^{g\rho}, \quad (21)$$

which asserts that gM is defined in such a way that measuring it on the transformed state $g\rho$ just gives the translated probability $g(M^\rho)$. Note that the parenthesis order actually matters in (21).

A fixed point M for the action of G on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is a *G-covariant observable*, i.e. $U(g)M(x, y)U(g)^* = M(g(x, y))$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $g \in G$. On the other hand, if $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is any observable, then

$$M_G = \frac{1}{|G|} \sum_{g \in G} gM \quad (22)$$

is a *G-covariant* element in $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$, which we call the *covariant version* of M .

Now we state some sufficient conditions on the observables A, B and on the action of the group G ensuring that the entropic divergence $D(A, B \| \cdot)$ is *G-invariant*, and then we derive their consequences on the optimal approximate joint measurements of A and B .

Note that the relative entropy is always invariant for a group action, that is,

$$S(gp \| gq) = S(p \| q), \quad \forall p, q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), \quad g \in G, \quad (23)$$

by item (iv) of Proposition 1. Note also that, for $p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, the expression $gp_{[i]} = (gp)_{[i]}$ is unambiguous, as the action of g is defined on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and not on $\mathcal{P}(\mathcal{X})$ or $\mathcal{P}(\mathcal{Y})$.

Theorem 8. *Let $A \in \mathcal{M}(\mathcal{X})$, $B \in \mathcal{M}(\mathcal{Y})$ be the reference observables. Let G be a finite group, acting on $\mathcal{X} \times \mathcal{Y}$ and with a projective unitary representation U on \mathcal{H} . Suppose the group G is generated by a subset $S_G \subseteq G$, such that each $g \in S_G$ satisfies either one condition between:*

- (i) *there exist maps $f_{g,\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{X}$ and $f_{g,\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{Y}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,*
 - (a) $g(x, y) = (f_{g,\mathcal{X}}(x), f_{g,\mathcal{Y}}(y))$
 - (b) $U(g)A(x)U(g)^* = A(f_{g,\mathcal{X}}(x))$ and $U(g)B(y)U(g)^* = B(f_{g,\mathcal{Y}}(y))$;
- (ii) *there exist maps $f_{g,\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{Y}$ and $f_{g,\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{X}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,*
 - (a) $g(x, y) = (f_{g,\mathcal{Y}}(y), f_{g,\mathcal{X}}(x))$

$$(b) \ U(g)A(x)U(g)^* = B(f_{g,x}(x)) \text{ and } U(g)B(y)U(g)^* = A(f_{g,y}(y)).$$

Then, $D(A, B \| gM) = D(A, B \| M)$ for all $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ and $g \in G$.

Proof. If two elements $g_1, g_2 \in G$ satisfy the above hypotheses, so does their product $g_1 g_2$. Since S_G generates G , we can then assume that $S_G = G$. In this case, condition (i.a) or (ii.a) easily implies the relation

$$gp_{[1]} \otimes gp_{[2]} = g(p_{[1]} \otimes p_{[2]}) \quad \forall p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), g \in G. \quad (24)$$

On the other hand, by condition (i.b) or (ii.b), we get

$$A^{g\rho} \otimes B^{g\rho} = g(A^\rho \otimes B^\rho) \quad \forall \rho \in \mathcal{S}(\mathcal{H}), g \in G. \quad (25)$$

For any $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we then have

$$\begin{aligned} S[A, B \| g^{-1}M](\rho) &= S(A^\rho \otimes B^\rho \| (g^{-1}M)_{[1]}^\rho \otimes (g^{-1}M)_{[2]}^\rho) && \text{by (5)} \\ &= S(A^\rho \otimes B^\rho \| g^{-1}(M^{g\rho})_{[1]} \otimes g^{-1}(M^{g\rho})_{[2]}) && \text{by (21)} \\ &= S(A^\rho \otimes B^\rho \| g^{-1}(M_{[1]}^{g\rho} \otimes M_{[2]}^{g\rho})) && \text{by (24)} \\ &= S(g(A^\rho \otimes B^\rho) \| M_{[1]}^{g\rho} \otimes M_{[2]}^{g\rho}) && \text{by (23)} \\ &= S(A^{g\rho} \otimes B^{g\rho} \| M_{[1]}^{g\rho} \otimes M_{[2]}^{g\rho}) && \text{by (25)} \\ &= S[A, B \| M](g\rho). \end{aligned}$$

Taking the supremum over ρ and observing that $\mathcal{S}(\mathcal{H}) = g\mathcal{S}(\mathcal{H})$, it follows that $D(A, B \| g^{-1}M) = D(A, B \| M)$. \square

- Remark 4.**
1. The occurrence of either hypothesis (i) or (ii) may depend on the generator $g \in S_G$, and needs not to be the same for the whole S_G .
 2. Conditions (i.a), (ii.a) are hypotheses about the action of G on the outcome space $\mathcal{X} \times \mathcal{Y}$. Note that each one implies that the maps $f_{g,x}$ and $f_{g,y}$ are bijective. In particular, one can have some generators g satisfying (ii.a) only if $|\mathcal{X}| = |\mathcal{Y}|$.
 3. Conditions (i.b), (ii.b) involve also the observables A and B . Even if A and B are not compatible, they are asked to behave as if they were the marginals of a covariant bi-observable.
 4. The symmetries allowed in hypothesis (ii) of Theorem 8 essentially are of permutational nature. They directly follow from the exchange symmetry of the error function (4), in which the approximation errors $S(A^\rho \| M_{[1]}^\rho)$ and $S(B^\rho \| M_{[2]}^\rho)$ are equally weighted.

Corollary 9. *Under the hypotheses of Theorem 8,*

- the set $\mathcal{M}_{\text{inc}}(A, B)$ is G -invariant;
- for any $M \in \mathcal{M}_{\text{inc}}(A, B)$, we have $M_G \in \mathcal{M}_{\text{inc}}(A, B)$;
- there exists a G -covariant observable in $\mathcal{M}_{\text{inc}}(A, B)$.

Proof. Since $D(A, B \| \cdot)$ is G -invariant by Theorem 8, then the set $\mathcal{M}_{\text{inc}}(A, B)$ is G -invariant. This fact and the convexity of $\mathcal{M}_{\text{inc}}(A, B)$ implies that $M_G \in \mathcal{M}_{\text{inc}}(A, B)$ for all $M \in \mathcal{M}_{\text{inc}}(A, B)$. Since the latter set is nonempty by Theorem 6, item (iv), it then always contains a G -covariant observable. \square

Remark 5. Since the covariance requirement reduces the many degrees of freedom in the choice of a bi-observable $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, we expect that the larger is the symmetry group G , the minor amount of free parameters will be needed to describe a G -covariant element in $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$. This will be a big help in the computation of $c_{\text{inc}}(A, B)$, as Corollary 9 allows to minimize $D(A, B \| \cdot)$ just on the set of G -covariant bi-observables. More precisely, under the hypotheses of Theorem 8,

$$c_{\text{inc}}(A, B) = \min_{\substack{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \\ M \text{ } G\text{-covariant}}} \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \left\{ S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho) \right\},$$

where the minimum has to be computed only with respect to the parameters describing a G -covariant bi-observable M . In particular, it is only the dependence of the marginals $M_{[1]}$ and $M_{[2]}$ on such parameters that comes into play. Of course, solving this double optimization problem yields the value of $c_{\text{inc}}(A, B)$ and all the covariant optimal joint measurement of A and B , but not the whole optimal set $\mathcal{M}_{\text{inc}}(A, B)$.

In the cases of two othogonal spin components (Section 3.2.1) and of a pair of MUB (Section 3.3), covariance will reduce the number of parameters to just a single one.

If B is not sharp, the two sets $\mathcal{M}_{\text{inc}}(A, B)$ and $\mathcal{M}_{\text{ed}}(A, B)$ may be different, and we need a specific corollary for $\mathcal{M}_{\text{ed}}(A, B)$. Indeed, stronger hypotheses are required to ensure that the set $\mathcal{M}(\mathcal{X}; B)$ of the sequential measurements is G -invariant.

Corollary 10. *Under the hypotheses of Theorem 8, and supposing in addition that all the generators $g \in S_G$ enjoy only condition (i),*

- *the set $\mathcal{M}(\mathcal{X}; B)$ is G -invariant;*
- *the set $\mathcal{M}_{\text{ed}}(A, B)$ is G -invariant;*
- *for any $M \in \mathcal{M}_{\text{ed}}(A, B)$, we have $M_G \in \mathcal{M}_{\text{ed}}(A, B)$;*
- *there exists a G -covariant observable in $\mathcal{M}_{\text{ed}}(A, B)$.*

Proof. We know that $D(A, B \| \cdot)$ is G -invariant by Theorem 8, and so we only have to prove that $\mathcal{M}(\mathcal{X}; B)$ is G -invariant; then the subsequent theses follow as in Corollary 9. Since we can assume $S_G = G$, any element $g \in G$ maps a sequential measurement $M = \mathcal{J}^*(B)$ to another sequential measurement $\mathcal{J}'^*(B)$, due to condition (i) of Theorem 8:

$$\begin{aligned} gM(x, y) &= U(g)M\left(g^{-1}(x, y)\right)U(g)^* = U(g)M\left(f_{g, \mathcal{X}}^{-1}(x), f_{g, \mathcal{Y}}^{-1}(y)\right)U(g)^* \\ &= U(g)\mathcal{J}_{f_{g, \mathcal{X}}^{-1}(x)}^*\left[B\left(f_{g, \mathcal{Y}}^{-1}(y)\right)\right]U(g)^* = U(g)\mathcal{J}_{f_{g, \mathcal{X}}^{-1}(x)}^*\left[U(g)^*B(y)U(g)\right]U(g)^* := \mathcal{J}'^*(B)(x, y). \end{aligned}$$

□

Remark 6. Corollary 10 does not admit elements g satisfying condition (ii) of Theorem 8 because this hypothesis alone can not guarantee the G -invariance of the set $\mathcal{M}(\mathcal{X}; B)$. Of course, it works for a sharp B , but it could fail, for example, for a trivial B . Indeed, take $\mathcal{X} = \mathcal{Y}$ and $A = B = U_{\mathcal{X}}$; then $\mathcal{M}(\mathcal{X}; B) = \{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) : M(x, y) = M_1(x)u_{\mathcal{X}}(y), \forall x, y, \text{ for some } M_1 \in \mathcal{M}(\mathcal{X})\}$, and $M_{[2]}(y) = B(y)$ has rank d for every $M \in \mathcal{M}(\mathcal{X}; B)$ and $y \in \mathcal{Y}$. Nevertheless, if g satisfies (ii.a), then (ii.b) is obvious, but g could map a sequential measurement M outside $\mathcal{M}(\mathcal{X}; B)$. Indeed, $(gM)_{[2]}(y) = U(g)M_1\left(f_{g, \mathcal{X}}^{-1}(y)\right)U(g)^*$ has rank equal to the rank of $M_1\left(f_{g, \mathcal{X}}^{-1}(y)\right)$, which can be chosen smaller than d .

3.2 Two spin-1/2 components

As a first application of Theorem 8 and its corollaries, we take as the reference observables two spin-1/2 components along the directions defined by two unit vectors \mathbf{a} and \mathbf{b} in \mathbb{R}^3 . They are represented by the sharp observables

$$A(x) = \frac{1}{2}(\mathbb{1} + x\mathbf{a} \cdot \boldsymbol{\sigma}), \quad B(y) = \frac{1}{2}(\mathbb{1} + y\mathbf{b} \cdot \boldsymbol{\sigma}), \quad (26)$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of the three Pauli matrices on $\mathcal{H} = \mathbb{C}^2$, and $\mathcal{X} = \mathcal{Y} = \{-1, +1\}$. We consider the general case of incompatible A and B ; that is, \mathbf{a} and \mathbf{b} may be any couple of non parallel directions. Let $\alpha \in (0, \pi)$ be the angle formed by \mathbf{a} and \mathbf{b} ; by item (i) in Theorem 6, the coefficient $c_{\text{inc}}(A, B)$ does not depend on the choice of the values of the outcomes, and this allows us to take $\alpha \in (0, \pi/2]$. Indeed, when $\alpha > \pi/2$, it is enough to change $y \rightarrow -y$ and $\mathbf{b} \rightarrow -\mathbf{b}$ to recover the case $\alpha \in (0, \pi/2]$. Without loss of generality, we take the two spin directions in the $\mathbf{i}\mathbf{j}$ -plane and choose the \mathbf{i} -

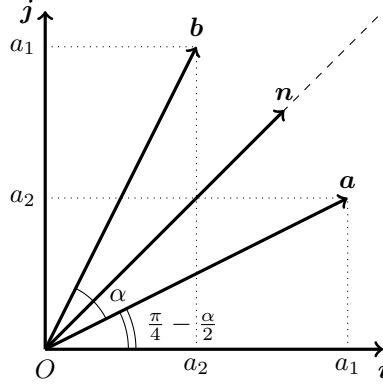


Figure 1: The unit vectors \mathbf{a} and \mathbf{b} characterizing the sharp spin-1/2 observables (26)

and \mathbf{j} -axes in such a way that the bisector of the angle formed by \mathbf{a} and \mathbf{b} coincides with the bisector \mathbf{n} of the first quadrant. This choice is illustrated in Figure 1, where

$$\begin{aligned} \mathbf{a} &= a_1 \mathbf{i} + a_2 \mathbf{j}, & \mathbf{b} &= a_2 \mathbf{i} + a_1 \mathbf{j}, & \mathbf{n} &= \frac{\mathbf{i} + \mathbf{j}}{\sqrt{2}}, & a_1^2 + a_2^2 &= 1, \\ \alpha &\in \left(0, \frac{\pi}{2}\right], & a_1 &= \sqrt{\frac{1 + \sin \alpha}{2}} \in \left(\frac{1}{\sqrt{2}}, 1\right], & a_2 &= \frac{\cos \alpha}{\sqrt{2(1 + \sin \alpha)}} \in \left[0, \frac{1}{\sqrt{2}}\right). \end{aligned} \quad (27)$$

3.2.1 Covariance properties

We now analyse the covariance properties of the two spin observables A and B , and study the general structure of the corresponding covariant bi-observables.

Let $D_2 \subset SO(3)$ be the order 4 dihedral group generated by the rotations $S_{D_2} = \{R_{\mathbf{n}}(\pi), R_{\mathbf{m}}(\pi)\}$, i.e., the 180° rotations around the bisector of the first quadrant \mathbf{n} and the bisector of the second quadrant $\mathbf{m} = (\mathbf{j} - \mathbf{i})/\sqrt{2}$ (here and in the following, our reference for the discrete subgroups of the rotation group is [51, pp. 77–79]). The natural action of the group D_2 on the outcome space $\mathcal{X} \times \mathcal{Y}$ is given by

$$R_{\mathbf{n}}(\pi)(x, y) = (y, x) \quad \text{and} \quad R_{\mathbf{m}}(\pi)(x, y) = (-y, -x), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (28a)$$

We then see that condition (ii.a) of Theorem 8 is satisfied for all $g \in S_{D_2}$. As the representation U of the symmetry group on \mathbb{C}^2 , we take the restriction to D_2 of the usual spin-1/2 projective representation of $SO(3)$; this gives

$$U(R_{\mathbf{n}}(\pi)) = e^{-i\pi \mathbf{n} \cdot \boldsymbol{\sigma}/2} \equiv -i \mathbf{n} \cdot \boldsymbol{\sigma}, \quad U(R_{\mathbf{m}}(\pi)) = e^{-i\pi \mathbf{m} \cdot \boldsymbol{\sigma}/2} \equiv -i \mathbf{m} \cdot \boldsymbol{\sigma}. \quad (28b)$$

It is easy to see that the observables A and B satisfy the relations

$$\begin{aligned} U(R_{\mathbf{n}}(\pi))A(x)U(R_{\mathbf{n}}(\pi))^* &= B(x), & U(R_{\mathbf{n}}(\pi))B(y)U(R_{\mathbf{n}}(\pi))^* &= A(y), \\ U(R_{\mathbf{m}}(\pi))A(x)U(R_{\mathbf{m}}(\pi))^* &= B(-x), & U(R_{\mathbf{m}}(\pi))B(y)U(R_{\mathbf{m}}(\pi))^* &= A(-y). \end{aligned} \quad (29)$$

This implies that also condition (ii.b) of Theorem 8 is satisfied for all $g \in S_{D_2}$. Then, because of Remark 5, in order to find $c_{\text{inc}}(A, B)$, we are led to study the most general form of a D_2 -covariant bi-observable and its marginals.

Proposition 11. *Let the dihedral group D_2 act on $\mathcal{X} \times \mathcal{Y}$ and \mathcal{H} as in (28). Then the following facts hold:*

(i) *The most general D_2 -covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$ is*

$$\mathbf{M}(x, y) = \frac{1}{4} [(1 + \gamma xy) \mathbb{1} + (c_1 x + c_2 y) \sigma_1 + (c_2 x + c_1 y) \sigma_2], \quad (30)$$

with $\gamma, c_1, c_2 \in \mathbb{R}$ such that

$$\sqrt{2}|c_1 + c_2| - 1 \leq \gamma \leq 1 - \sqrt{2}|c_1 - c_2|. \quad (31)$$

The marginals of M are $M_{[1]} = A_c$ and $M_{[2]} = B_c$, where

$$A_c(x) = \frac{1}{2} [\mathbb{1} + x(c_1\sigma_1 + c_2\sigma_2)], \quad B_c(y) = \frac{1}{2} [\mathbb{1} + y(c_2\sigma_1 + c_1\sigma_2)], \quad c = c_1\mathbf{i} + c_2\mathbf{j}. \quad (32)$$

(ii) Equation (32) defines the marginals of a D_2 -covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$ if and only if

$$|c_1| \leq 1/\sqrt{2}, \quad |c_2| \leq 1/\sqrt{2}. \quad (33)$$

Note that the components of c appear in A_c and B_c in the reverse order.

Proof. (i) The set $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is a subset of the linear space $\mathcal{L}(\mathbb{C}^2)^{\mathcal{X} \times \mathcal{Y}} = \mathbb{C}^{\mathcal{X} \times \mathcal{Y}} \otimes \mathcal{L}(\mathbb{C}^2)$, where the set of the 16 products between one of the functions $1, x, y, xy$ with one of the operators $\mathbb{1}, \sigma_1, \sigma_2, \sigma_3$ provides a basis of linearly independent elements. Then, the structure of the most general bi-observable on $\mathcal{X} \times \mathcal{Y}$ is a linear combination of such products; it is easy to see that the covariance under the rotation $R_n(\pi)R_m(\pi)$ implies the vanishing of the coefficients of the products $x\mathbb{1}, y\mathbb{1}, xy\sigma_1, xy\sigma_2, 1\sigma_1, 1\sigma_2, x\sigma_3, y\sigma_3$. By taking into account also the normalization and selfadjointness conditions, we are left with

$$M(x, y) = \frac{1}{4} [(1 + \gamma xy) \mathbb{1} + (c_1x + c_2y) \sigma_1 + (c'_1x + c'_2y) \sigma_2 + (c_3 + c_4xy) \sigma_3],$$

with real coefficients γ, c_i and c'_i . By imposing the covariance of this expression under $R_n(\pi)$ we get $c'_1 = c_2, c'_2 = c_1, c_3 = c_4 = 0$, and (30) is obtained. Finally, one can check that (30) is covariant also under $R_m(\pi)$, hence with respect to the whole group D_2 . To impose the positivity of the operators $M(x, y)$ it is enough to study the diagonal elements and the determinant of the 2×2 -matrix representing (30). The positivity of the diagonal elements $\forall(x, y)$ gives $\gamma \in [-1, 1]$. This condition and the positivity of the determinant, i.e.,

$$(1 + \gamma xy)^2 \geq (c_1x + c_2y)^2 + (c_2x + c_1y)^2, \quad \forall(x, y) \in \mathcal{X} \times \mathcal{Y},$$

altogether are equivalent to (31). Evaluating the marginals of (30) immediately yields (32).

(ii) We begin by proving that inequalities (33) are equivalent to

$$\sqrt{2}|c_1 + c_2| - 1 \leq 1 - \sqrt{2}|c_1 - c_2|, \quad (34)$$

Indeed, (34) rewrites $|c_1 + c_2| + |c_1 - c_2| \leq \sqrt{2}$; by checking all the possible sign occurrences in c_1 and c_2 , the latter inequality is equivalent to $|c_1| + |c_2| + ||c_1| - |c_2|| \leq \sqrt{2}$. In turn, by separately inspecting the two cases $|c_1| \geq |c_2|$ and $|c_2| \geq |c_1|$ this is the same as (33).

Now, for the marginals A_c and B_c of a D_2 -covariant bi-observable, inequalities (31) trivially imply (34), and so (33). Conversely, if A_c and B_c are as in (32) with c_1, c_2 satisfying (33), then by (34) we can always find γ as in (31). The D_2 -covariant bi-observable corresponding to γ, c_1, c_2 then has marginals A_c and B_c . \square

The compatibility condition (33) corresponds to the vector c lying in the square

$$Q = \{c_1\mathbf{i} + c_2\mathbf{j} : |c_1| \leq 1/\sqrt{2}, |c_2| \leq 1/\sqrt{2}\}. \quad (35)$$

On the other hand, when one drops any requirement about compatibility, formula (32) defines two observables A_c and B_c if and only if $|c| \leq 1$, that is, c belongs to the disk

$$C = \{c_1\mathbf{i} + c_2\mathbf{j} : c_1^2 + c_2^2 \leq 1\}. \quad (36)$$

Note that, for $c \neq 0$, the observable A_c is a noisy version of the sharp observable $A_{c/|c|}$, which is the spin-1/2 component along the direction $c/|c|$:

$$A_c = |c| A_{\frac{c}{|c|}} + (1 - |c|)U_{\mathcal{X}}$$

(cf. (14) with $\lambda = |c|$). Analogue considerations hold for B_c .

Orthogonal components. When the angle between the spin directions \mathbf{a} and \mathbf{b} is $\alpha = \pi/2$, the reference observables become the two orthogonal spin components along the \mathbf{i} - and \mathbf{j} -axes:

$$A(x) = X(x) := \frac{1}{2} (\mathbb{1} + x\sigma_1), \quad B(y) = Y(y) := \frac{1}{2} (\mathbb{1} + y\sigma_2). \quad (37)$$

In this case, the symmetries of our system increase from D_2 to the enlarged dihedral group D_4 . Here we recall that $D_4 \subset SO(3)$ is the order 8 group of the 90° rotations around the \mathbf{k} -axis, together with the 180° rotations around \mathbf{i} , \mathbf{j} , \mathbf{n} and \mathbf{m} ; clearly, $D_2 \subset D_4$. Now, the two rotations $S_{D_4} = \{R_{\mathbf{i}}(\pi), R_{\mathbf{n}}(\pi)\}$ generate D_4 ; for instance, we have $R_{\mathbf{j}}(\pi) = R_{\mathbf{n}}(\pi)R_{\mathbf{i}}(\pi)R_{\mathbf{n}}(\pi)$, $R_{\mathbf{m}}(\pi) = R_{\mathbf{i}}(\pi)R_{\mathbf{n}}(\pi)R_{\mathbf{i}}(\pi)$, $R_{\mathbf{k}}(\pi/2) = R_{\mathbf{m}}(\pi)R_{\mathbf{j}}(\pi)$.

The action of the group element $R_{\mathbf{n}}(\pi)$ on $\mathcal{X} \times \mathcal{Y}$, $A = X$ and $B = Y$ is still given by (28a) and (29), respectively; we have already seen that these actions satisfy condition (ii) of Theorem 8. Further, by introducing the natural actions

$$R_{\mathbf{i}}(\pi)(x, y) = (x, -y), \quad U(R_{\mathbf{i}}(\pi)) = e^{-i\pi \mathbf{i} \cdot \boldsymbol{\sigma}/2} \equiv -i \mathbf{i} \cdot \boldsymbol{\sigma}, \quad (38)$$

we have

$$U(R_{\mathbf{i}}(\pi))X(x)U(R_{\mathbf{i}}(\pi))^* = X(x), \quad U(R_{\mathbf{i}}(\pi))Y(y)U(R_{\mathbf{i}}(\pi))^* = Y(-y). \quad (39)$$

In particular, we see that $R_{\mathbf{i}}(\pi)$ fulfills condition (i) of the same theorem. Therefore, all $g \in S_{D_4}$ satisfy the hypotheses of Theorem 8.

Again, by Remark 5, we now look for the general expression of a D_4 -covariant bi-observable and its marginals.

Corollary 12. *Let the dihedral group D_4 act on $\mathcal{X} \times \mathcal{Y}$ and on \mathcal{H} by (28) and (38). Then, the most general D_4 -covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$ is given by (30) with $\gamma = 0$, $c_2 = 0$ and $|c_1| \leq 1/\sqrt{2}$, that is,*

$$M(x, y) = \frac{1}{4} [\mathbb{1} + c_1 (x\sigma_1 + y\sigma_2)], \quad |c_1| \leq 1/\sqrt{2}. \quad (40)$$

Proof. By applying the extra transformation (38) to the D_2 -covariant observable (30) we get

$$\begin{aligned} R_{\mathbf{i}}(\pi)M(x, y) &= U(R_{\mathbf{i}}(\pi))M(x, -y)U(R_{\mathbf{i}}(\pi))^* \\ &= \frac{1}{4} [(1 - \gamma xy) \mathbb{1} + (c_1 x - c_2 y) \sigma_1 - (c_2 x - c_1 y) \sigma_2]. \end{aligned}$$

In order to have covariance also under this transformation, it must be $\gamma = 0$ and $c_2 = 0$; then, conditions (33) become the inequality in (40). \square

3.2.2 Incompatibility degree for two spin components

Now we search for the value of the lower bound $c_{\text{inc}}(A, B)$ and the optimal covariant approximate joint measurement in $\mathcal{M}_{\text{inc}}(A, B)$. By Remark 5 and Proposition 11,

$$\begin{aligned} c_{\text{inc}}(A, B) &= \min_{\substack{M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \\ M \text{ } D_2\text{-covariant}}} \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \{S(A^\rho \| M_{[1]}^\rho) + S(B^\rho \| M_{[2]}^\rho)\} \\ &= \min_{c \in Q} \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \{S(A^\rho \| A_c^\rho) + S(B^\rho \| B_c^\rho)\}, \end{aligned} \quad (41)$$

where Q is the square (35). Thus, the value of $c_{\text{inc}}(A, B)$ can be found by minimizing the function

$$D(c) = \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \{S(A^\rho \| A_c^\rho) + S(B^\rho \| B_c^\rho)\} \quad (42)$$

for c ranging inside Q .

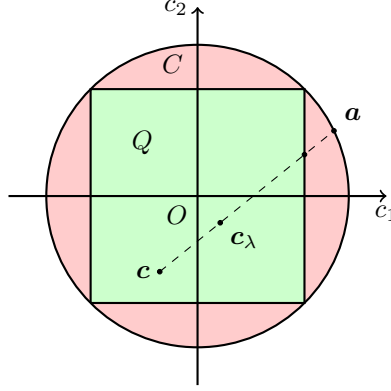


Figure 2: The domain of the function D defined in (42) (the disk C) and the set of c over which D is minimized (the square Q).

Note that the domain of the function D can be extended to the whole disk C introduced in (36). In the domain C , $D(c) = 0$ if and only if $A_c = A$ and $B_c = B$, that is equivalent to $c = a$. The regions C and Q in the \mathbf{ij} -plane are depicted in Figure 2.

By using convexity arguments, we now prove that the minimization of the function D over the square Q fixes $c_1 = 1/\sqrt{2}$. This considerably simplifies the search of an optimal D_2 -covariant bi-observable, as it reduces the involved parameters from the number of three to a single one.

Proposition 13. *Let A and B be the spin-1/2 reference observables (26). Let the dihedral group D_2 act on $\mathcal{X} \times \mathcal{Y}$ and \mathcal{H} as in (28). Then, the most general D_2 -covariant bi-observable in the set $\mathcal{M}_{\text{inc}}(A, B)$ is of the form*

$$M_\gamma(x, y) = \frac{1}{4} \left[(1 + \gamma xy) \mathbb{1} + \frac{1}{\sqrt{2}} (x\sigma_1 + y\sigma_2) + \frac{\gamma}{\sqrt{2}} (y\sigma_1 + x\sigma_2) \right] \quad (43)$$

with $\gamma = \sqrt{2}c_2 \in [-1, 1]$ and c_2 determined by the minimization problem

$$c_{\text{inc}}(A, B) = \min_{|c_2| \leq 1/\sqrt{2}} D \left(\frac{1}{\sqrt{2}} \mathbf{i} + c_2 \mathbf{j} \right). \quad (44)$$

Proof. The mappings $c \mapsto A_c^\rho$ and $c \mapsto B_c^\rho$ are affine on the disk C for all $\rho \in \mathcal{S}(\mathcal{H})$, which, together with the convexity of the relative entropy, implies that also the mappings $c \mapsto S(A^\rho \| A_c^\rho)$ and $c \mapsto S(B^\rho \| B_c^\rho)$ are convex; hence, such are their sum and the supremum (42). Moreover, we already noticed that $D(c) = 0$ if and only if $c = a$.

Now, by (41), we have

$$c_{\text{inc}}(A, B) = \min_{c \in Q} D(c).$$

Making reference to Figure 2, let us take $c \in Q$ and introduce the line segment joining c and a : $c_\lambda = (1 - \lambda)c + \lambda a$, $\lambda \in [0, 1]$. By defining $D(\lambda) = D(c_\lambda)$ we get that $\lambda \mapsto D(\lambda)$ is a strictly decreasing convex function on $[0, 1]$ with $D(1) = D(a) = 0$. Indeed, for $0 \leq \lambda' < \lambda < 1$, one has

$$D(\lambda) = D \left(\frac{1 - \lambda}{1 - \lambda'} c_{\lambda'} + \frac{\lambda - \lambda'}{1 - \lambda'} a \right) \leq \frac{1 - \lambda}{1 - \lambda'} D(c_{\lambda'}) + \frac{\lambda - \lambda'}{1 - \lambda'} D(a) < D(\lambda').$$

Then, the minimum of $D(c_\lambda)$ with respect to $c_\lambda \in Q$ is where the line segment crosses the right side of the square, i.e. for $(c_\lambda)_1 = 1/\sqrt{2}$. This is true for every point c in the square Q . Therefore, the minimum points of $D(c)$ for $c \in Q$ need to be on the right edge $\{1/\sqrt{2}\mathbf{i} + c_2\mathbf{j} : |c_2| \leq 1/\sqrt{2}\}$ of Q itself.

Now, let M be a D_2 -covariant bi-observable on $\mathcal{X} \times \mathcal{Y}$, that we parameterize with γ , c_1 and c_2 as in (30). Then its marginals are $M_{[1]} = A_c$ and $M_{[2]} = B_c$, with $c \in Q$ by (33). Therefore,

$$D(A, B \| M) = D(c), \quad c \in Q.$$

If $M \in \mathcal{M}_{\text{inc}}(A, B)$, then $c_1 = 1/\sqrt{2}$ and c_2 is a solution of the minimization problem (44). Indeed, if this were not the case, we would have $D(A, B \| M) = D(c) > c_{\text{inc}}(A, B)$ by the previous paragraph. Finally, the constraint $c_1 = 1/\sqrt{2}$ and (31) fix $\gamma = \sqrt{2}c_2$. \square

It is worth noticing that every D_2 -covariant bi-observable (43) can be rewritten as a *mixture* (convex combination) of two sharp joint measurements of compatible spin components, along the bisector \mathbf{n} in the case of the first bi-observable and along the bisector \mathbf{m} for the other one. More precisely, we introduce the sharp bi-observables

$$\begin{aligned} M_+(x, y) &= \left[\frac{1}{2}(\mathbb{1} + x\mathbf{n} \cdot \boldsymbol{\sigma}) \right] \left[\frac{1}{2}(\mathbb{1} + y\mathbf{n} \cdot \boldsymbol{\sigma}) \right] \equiv A_{\mathbf{n}}(x)B_{\mathbf{n}}(y), \\ M_-(x, y) &= \left[\frac{1}{2}(\mathbb{1} - x\mathbf{m} \cdot \boldsymbol{\sigma}) \right] \left[\frac{1}{2}(\mathbb{1} + y\mathbf{m} \cdot \boldsymbol{\sigma}) \right] \equiv A_{-\mathbf{m}}(x)B_{-\mathbf{m}}(y), \end{aligned} \quad (45)$$

where A_c, B_c are defined by (32) (which implies $B_{\mathbf{n}} = A_{\mathbf{n}}, B_{-\mathbf{m}} = A_{\mathbf{m}}$); then, we have

$$M_\gamma = \frac{1+\gamma}{2} M_+ + \frac{1-\gamma}{2} M_-. \quad (46)$$

The bi-observable M_γ can be expressed also as the alternative mixture

$$M_\gamma = \begin{cases} \gamma M_+ + (1-\gamma)M_0 & \text{if } \gamma = \sqrt{2}c_2 \geq 0, \\ |\gamma| M_- + (1-|\gamma|)M_0 & \text{if } \gamma = \sqrt{2}c_2 \leq 0. \end{cases}$$

where the bi-observable M_0 , defined by (47) below, will play a special role in the orthogonal case.

Orthogonal components. For the orthogonal spin components X and Y , there is at least a D_4 -covariant bi-observable $M \in \mathcal{M}_{\text{inc}}(X, Y)$. This bi-observable is clearly also D_2 -covariant, hence, comparing (40) and (43), it follows that

$$M_0(x, y) = \frac{1}{4} \left(\mathbb{1} + \frac{x}{\sqrt{2}} \sigma_1 + \frac{y}{\sqrt{2}} \sigma_2 \right) \quad (47)$$

is the unique D_4 -covariant bi-observable in $\mathcal{M}_{\text{inc}}(X, Y)$. Note that $M_0(x, y)$ is a rank-one operator for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Using the sharp spin measurements (45), the optimal approximate joint measurement M_0 can be rewritten as the mixture

$$M_0 = \frac{1}{2}(M_+ + M_-).$$

In the orthogonal case we have $\mathbf{a} = \mathbf{i}$ and $\mathbf{c} = \mathbf{i}/\sqrt{2} \parallel \mathbf{a}$, so that the marginals of the optimally approximating joint measurement are noisy versions of the reference observables.

By optimality, we have that $c_{\text{inc}}(X, Y) = D(X, Y \| M_0)$. In Appendix B.1 we solve also the maximization problem in the evaluation of $D(X, Y \| M_0)$, and show that $D(X, Y \| M_0) = S[X, Y \| M_0](\rho_e)$, where ρ_e is the projection onto any of the eigenvectors of σ_1 or σ_2 . The entropic incompatibility degree then takes the value

$$c_{\text{inc}}(X, Y) = \log \left[2 \left(2 - \sqrt{2} \right) \right] \simeq 0.228447. \quad (48)$$

Remarkably, the knowledge of the value of $c_{\text{inc}}(X, Y)$ allows us to prove that M_0 is not only the unique optimal D_4 -covariant joint measurement of X and Y , but actually it is the unique element in the whole set $\mathcal{M}_{\text{inc}}(X, Y)$.

Proposition 14. *Let X and Y be the two orthogonal spin-1/2 components along the \mathbf{i} - and \mathbf{j} -axis introduced in (37). Then*

$$\mathcal{M}_{\text{inc}}(X, Y) = \{M_0\}, \quad (49)$$

where M_0 is the bi-observable (47).

Proof. Let M be a bi-observable in $\mathcal{M}_{\text{inc}}(X, Y)$. By Corollary 9, its covariant version M_{D_4} is still in $\mathcal{M}_{\text{inc}}(X, Y)$, and hence $M_{D_4} = M_0$ by the above discussion. Definition (22) implies $gM(x, y) \leq |D_4|M_{D_4}(x, y)$ for all g and x, y , hence in particular $M(x, y) \leq |D_4|M_{D_4}(x, y) = 8M_0(x, y)$ for all x, y . Since $M_0(x, y)$ has rank 1, it must then be

$$M(x, y) = f(x, y)M_0(x, y), \quad \forall x, y$$

for some nonnegative coefficients $f(x, y)$. Writing f in the linear basis $1, x, y, xy$ of $\mathbb{C}^{X \times Y}$, the normalization constraint $\sum_{x, y} M(x, y) = \sum_{x, y} f(x, y)M_0(x, y) = \mathbb{1}$ gives $f(x, y) = 1 + \epsilon xy$ for some real parameter ϵ . The positivity constraints $M(x, y) = f(x, y)M_0(x, y) \geq 0$ require $f(x, y) \geq 0$, which give $-1 \leq \epsilon \leq 1$.

Summing up, if $M \in \mathcal{M}_{\text{inc}}(X, Y)$, then $M(x, y) = (1 + \epsilon xy)M_0(x, y)$ for some $\epsilon \in [-1, 1]$. Let us show that the only possible parameter is $\epsilon = 0$. Indeed, the marginals of M are

$$M_{[1]}(x) = \frac{1}{2} \left[\mathbb{1} + \frac{x}{\sqrt{2}} (\sigma_1 + \epsilon \sigma_2) \right], \quad M_{[2]}(y) = \frac{1}{2} \left[\mathbb{1} + \frac{y}{\sqrt{2}} (\epsilon \sigma_1 + \sigma_2) \right],$$

whose distributions in the state $\rho_e = \frac{1}{2}(\mathbb{1} + \sigma_1)$ are

$$M_{[1]}^{\rho_e}(x) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{2}} \right), \quad M_{[2]}^{\rho_e}(y) = \frac{1}{2} \left(1 + \epsilon \frac{y}{\sqrt{2}} \right).$$

On the other hand, we have $X^{\rho_e}(x) = \delta_1(x)$ and $Y^{\rho_e}(y) = 1/2$, so that

$$\begin{aligned} c_{\text{inc}}(X, Y) &= D(X, Y \| M) \geq S[X, Y \| M](\rho_e) = S(X^{\rho_e} \| M_{[1]}^{\rho_e}) + S(Y^{\rho_e} \| M_{[2]}^{\rho_e}) \\ &= \log \frac{1}{\frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right)} + S(Y^{\rho_e} \| M_{[2]}^{\rho_e}) = c_{\text{inc}}(X, Y) + S(Y^{\rho_e} \| M_{[2]}^{\rho_e}), \end{aligned}$$

which implies $S(Y^{\rho_e} \| M_{[2]}^{\rho_e}) = 0$. Hence, $Y^{\rho_e} = M_{[2]}^{\rho_e}$, and $\epsilon = 0$ then follows. \square

Numerical and analytical results for the nonorthogonal components. In the case of two arbitrarily oriented spin components, the maximization/minimization problem (42) and (44), giving c_{inc} and the optimal c_2 , is hard to be solved analytically. In particular, it is not even clear whether the minimum point c_2 is unique or not. Nevertheless, the double optimization over the state ρ and over the parameter c_2 can be tackled numerically, and the resulting $c_{\text{inc}}(A, B)$ for 100 equally distant values α in the interval $(0, \pi/2]$ are plotted in Figure 3.

A good analytical lower bound for $c_{\text{inc}}(A, B)$ can be found by fixing a trial state ρ , considering the bi-observables M_γ of (43), where $\gamma = \sqrt{2}c_2$, and minimizing the error function $S[A, B \| M_\gamma](\rho)$ with respect to γ in $[-1, 1]$. Indeed, by the definition of the entropic divergence, $S[A, B \| M_\gamma](\rho) \leq D(A, B \| M_\gamma)$, and so $\min_\gamma S[A, B \| M_\gamma](\rho) \leq \min_\gamma D(A, B \| M_\gamma) = c_{\text{inc}}(A, B)$ by Proposition 13. A convenient choice for ρ , suggested by the results in the case of two orthogonal components, is to take any of the eigenstates ρ_e of $\mathbf{a} \cdot \boldsymbol{\sigma}$ or $\mathbf{b} \cdot \boldsymbol{\sigma}$, say the eigenstate of $\mathbf{a} \cdot \boldsymbol{\sigma}$ with positive eigenvalue. Then, we get

$$c_{\text{inc}}(A, B) \geq \min_\gamma S[A, B \| M_\gamma](\rho_e) =: LB(\alpha); \quad (50)$$

the explicit expression of $S[A, B \| M_\gamma](\rho_e)$ is given in (82). Due to the neater geometrical interpretation of $c_2 = \gamma/\sqrt{2}$ given in Figure 2, henceforth we will privilege the parametrization with c_2 rather than γ . In Appendix B.2, the minimum over c_2 is computed and, for $\alpha \neq \pi/2$, it is found in the point

$$c_2 = \frac{1}{a_1} \left(\ell - \frac{a_2}{\sqrt{2}} \right), \quad (51)$$

where

$$\ell = \frac{1}{2\sqrt{2}a_2} \left(\sqrt{u^2 + 8(1+u)a_2^2} - u \right), \quad (52)$$

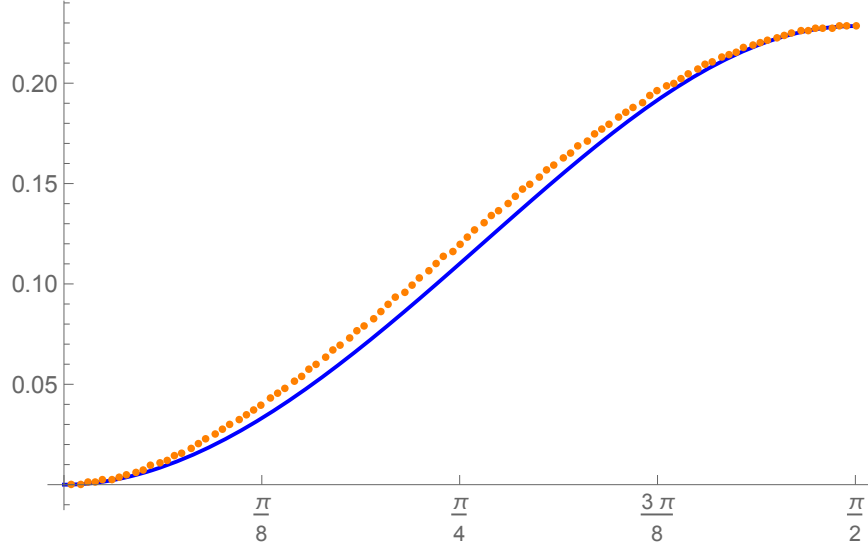


Figure 3: Dots: numerical evaluations of $c_{\text{inc}}(\mathbf{A}, \mathbf{B})$ as a function of α . Continuous line: the analytical lower bound $LB(\alpha)$.

$$u = \left(a_1 + \frac{1}{\sqrt{2}} \right) \frac{a_1^2 - a_2^2}{\sqrt{2}} = \left(1 + \sqrt{1 + \sin \alpha} \right) \frac{\sin \alpha}{2}. \quad (53)$$

In particular, the value (51) for c_2 , together with the fact that, for the D_2 -covariant bi-observable M_γ , the parameter $c_1 = 1/\sqrt{2}$ is fixed, show that the marginals of the bi-observable giving the lower bound (50) are not noisy versions of the reference observables \mathbf{A} and \mathbf{B} . Finally, the lower bound turns out to be

$$LB(\alpha) = -\log w + \frac{1}{2} (1 + \cos \alpha) \log \frac{1 + \cos \alpha}{1 + \ell} + \frac{1}{2} (1 - \cos \alpha) \log \frac{1 - \cos \alpha}{1 - \ell}, \quad (54)$$

with

$$w = \frac{1}{2} + \frac{\sqrt{u^2 + 8(1+u)a_2^2}}{4\sqrt{2}a_1} + \frac{\sin \alpha}{8} \left(\frac{3}{\sqrt{2}a_1} - 1 \right). \quad (55)$$

The plot of $LB(\alpha)$ is the continuous line in Figure 3.

For $\alpha = 0$, the reference observables are compatible and $c_{\text{inc}}(\mathbf{A}, \mathbf{B}) = 0$. For $\alpha \rightarrow 0$ the previous formulae give $u = 0$, $\ell = \pm 1$, $c_2 = a_2$, and one can check that also the lower bound (54) vanishes, as it must be (see also Remark 12 in Appendix B.2).

For two orthogonal components, i.e. $\alpha = \pi/2$, the expression (54) gives the exact value (48) of the entropic incompatibility degree, and it is not only a lower bound. This value can be computed by going to the limit $\alpha \rightarrow \pi/2$ in (54) or by using the result of Remark 13 in Appendix B.2.

For $\alpha \in (0, \pi/2)$, Figure 3 shows that the analytical lower bound (54) is not so far from the true value.

We now compare our optimal approximate joint measurements with other proposals coming from different approaches. Of course, every approximate joint measurement M that is optimal with respect to some other criterium will have a divergence from the target observables (\mathbf{A}, \mathbf{B}) larger or equal than $c_{\text{inc}}(\mathbf{A}, \mathbf{B})$. We stress that the two other proposals we will consider yield optimal bi-observables of the form M_γ , in which however the parameter γ (or, equivalently, c_2) is different from ours.

We have seen that, when $\alpha = \pi/2$, the incompatibility degree of \mathbf{A} and \mathbf{B} , as well as their unique optimal approximate joint measurement M_0 , can be evaluated analytically. In this special case, it turns out that M_0 is optimal also with respect to the other criteria we are going to consider in this section. However, this is not true for general α . In order to show it, we fix the angle $\alpha = \pi/4$, and compare the results of the different criteria in Table 1. We also add a LB column summarizing the parameters for the analytical lower bound (50). The rows provide: (1) the parameter $c_2 = \gamma/\sqrt{2}$ characterizing the measurement M_γ ; (2) the angle characterizing the pure state $\rho = \frac{1}{2} (\mathbb{1} + \cos \phi \sigma_x + \sin \phi \sigma_y)$ at which $S[\mathbf{A}, \mathbf{B} | M_\gamma](\rho)$ is computed,

that is the trial state ρ_e in the first column and the state maximizing $S[A, B \| M_\gamma]$ in the other ones; (3) the value of $S[A, B \| M_\gamma](\rho)$ for the parameters chosen in (1) and (2), which gives $LB(\pi/4)$ in the first column and the entropic divergence $D(A, B \| M_\gamma)$ in the other ones. The description of the columns is as follows.

Table 1: Incompatibility degree and its bounds for $\alpha = \pi/4$

criterium	LB	c_{inc}	BLW	NV
measurement: $c_2 \simeq$	0.562545	0.526087	0.382683	0.292893
state: $\phi \simeq$	0.392699	0.282743	0.391128	0.416889
value: $S[A, B \ M_\gamma](\rho) \simeq$	0.110081	0.120035	0.160886	0.212079

LB : The choice of the parameters is the one described in the computation of the analytical lower bound for c_{inc} . The parameter c_2 comes from (51), the angle $\phi = \pi/8$ corresponds to the trial state ρ_e (i.e., the eigenstate of $\mathbf{a} \cdot \boldsymbol{\sigma}$ for $\alpha = \pi/4$), and the corresponding value of $S[A, B \| M_\gamma](\rho)$ is the lower bound $LB(\pi/4)$.

c_{inc} : The parameters are chosen following the relative entropy approach to MUR. They are the numerical solution of the maximization/minimization problem (42), (44). Thus the value of $S[A, B \| M_\gamma](\rho)$ is the one found numerically for $c_{\text{inc}}(A, B)$, i.e., the dot at $\alpha = \pi/4$ in Figure 3; c_2 is the corresponding minimum point giving the optimal approximate joint measurement M_γ of A and B ; the angle ϕ corresponds to the state at which the error function $S[A, B \| M_\gamma]$ attains its maximum.

BLW: As discussed before Theorem 3, in [5–7, 11] a different approach is proposed. It is based on the Wasserstein distances between probabilities and involves two separate suprema over the states, so that the optimal approximate joint measurement is not unique. In [59], a best approximating measurement is obtained under a criterium which, using our notations, corresponds to the choice $c_2 = a_2$. Referring to Figure 2, this choice amounts to taking the point c inside the square Q as close as possible to \mathbf{a} , that is, letting c be the orthogonal projection of \mathbf{a} on the nearest square edge. The entropic divergence of the corresponding bi-observable M_γ from (A, B) and the state angle at which it is attained are the content of the BLW column.

NV: At the end of Section 2.3, we briefly discussed the proposal of [43] to use approximating joint measurements whose marginals are noisy versions (NV) of the two reference observables. In the present case, this corresponds to taking $c \parallel \mathbf{a}$, and the best choice is when c has the largest modulus, but is still inside the compatibility region Q ; that is, c is the intersection of the line joining \mathbf{a} and the origin with the right square edge. The results for this choice (together with the corresponding maximizing state) are reported in the last column.

3.3 Two conjugate observables in prime power dimension

We now consider two complementary observables in prime power dimension, realized by a couple of MUB that are conjugated by the Fourier transform of a finite field. The construction of a maximal set of MUB in a prime power dimensional Hilbert space by using finite fields is well known since Wootters and Fields' seminal paper [52]; see also [53, Sect. 2] for a review, and [54–56] for a group theoretical perspective on the topic. In this section, we will recall this construction in the simplest case of only two bases.

Let $\mathcal{X} = \mathcal{Y} \equiv \mathbb{F}$ be a finite field with characteristic p . We refer to [57, Sect. V.5] for the basic notions on finite fields. Here we only recall that p is a prime number, and \mathbb{F} has cardinality $|\mathbb{F}| = p^n$ for some positive integer n . We consider the $d = p^n$ -dimensional Hilbert space $\mathcal{H} = \ell^2(\mathbb{F})$, and choose as A and B the two observables

$$Q(x)\phi(z) = \delta_x(z)\phi(z) \quad \forall \phi \in \mathcal{H}, x, z \in \mathbb{F}, \quad (56)$$

and

$$P(y) = F^{-1}Q(y)F \quad \forall y \in \mathbb{F}, \quad (57)$$

in which $F : \mathcal{H} \rightarrow \mathcal{H}$ is the unitary *discrete Fourier transform*

$$F\phi(z) = \frac{1}{\sqrt{d}} \sum_{t \in \mathbb{F}} e^{-\frac{2\pi i}{p} \text{tr } zt} \phi(t) \quad (58)$$

and $\text{tr} : \mathbb{F} \rightarrow \mathbb{Z}_p$ is the field trace $\text{tr } x = \sum_{k=0}^{n-1} x^{p^k}$ (see [57, Sect. VI.5] for the definition and properties of the field trace). The observables Q and P are sharp, with $\text{rank } Q(x) = \text{rank } P(y) = 1$ for all x, y . More precisely, Q and P are the projection valued maps associated with the orthonormal bases $\{\delta_x\}_{x \in \mathbb{F}}$ and $\{\omega_y\}_{y \in \mathbb{F}}$, with $\omega_y(z) = (1/\sqrt{d}) e^{(2\pi i/p) \text{tr } yz}$, respectively. As $|\langle \delta_x | \omega_y \rangle| = 1/\sqrt{d}$ for all x and y , these two bases satisfy the MUB condition. In particular, as a consequence of [15], their preparation uncertainty coefficient is

$$c_{\text{prep}}(Q, P) = \log d. \quad (59)$$

Remark 7. The observables Q and P are an example of *Fourier conjugate MUB*, where in the present case the Fourier transform is taken with respect to the abelian additive group \mathbb{F} as in (58). This definition should be compared with the analogous one for MUB that are conjugated by means of the Fourier transform of the cyclic ring \mathbb{Z}_d , see e.g. [50]. In the latter case, the Hilbert space is $\mathcal{H} = \ell^2(\mathbb{Z}_d)$ and (58) is replaced by

$$\mathcal{F}\phi(z) = \frac{1}{\sqrt{d}} \sum_{t \in \mathbb{Z}} e^{-\frac{2\pi i}{d} zt} \phi(t)$$

(cf. [50, Eq. (4)]; note that no field trace appears in this formula). The two definitions are clearly the same if \mathbb{F} coincides with the cyclic field \mathbb{Z}_p (i.e. $n = 1$ and so $d = p$), but are essentially different in general. Indeed, as observed in [53, Sect. 5.3], they are inequivalent already for $d = 2^2$. The choice of using the field \mathbb{F} instead of the ring \mathbb{Z}_d in the present approach comes from the fact that MUB that are conjugated under the Fourier operator (58) actually share dilational symmetries that are not present in the \mathcal{F} -conjugate ones. Through the unitary representation D of the multiplicative group $\mathbb{F}_* := \mathbb{F} \setminus \{0\}$ defined below, these extra symmetries will allow us to turn the problem of characterizing the set $\mathcal{M}_{\text{inc}}(Q, P)$ into the optimization of a single parameter.

The natural symmetry group for the couple of observables (Q, P) is the group of the translations in the finite phase-space of the system together with all its symplectic transformations; as usual, we identify the latter symplectic group with the group $SL(2, \mathbb{F})$ of the 2×2 matrices with entries in \mathbb{F} and unit determinant. However, just a smaller subgroup of $SL(2, \mathbb{F})$ will be enough for us. Namely, for all $a \in \mathbb{F}_*$, we denote by $d(a)$ and $f(a)$ the $SL(2, \mathbb{F})$ -matrices

$$d(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad f(a) = \begin{pmatrix} 0 & a \\ -a^{-1} & 0 \end{pmatrix}.$$

Then the set $H = \{d(a), f(a) \mid a \in \mathbb{F}_*\}$ is an order $2(d-1)$ subgroup of the order $d(d^2-1)$ group $SL(2, \mathbb{F})$. It naturally acts by left multiplication on the additive group $V = \mathbb{F}^2$ of the \mathbb{F} -valued 2-entries column vectors $\mathbf{u} = (u_1, u_2)^T$. We can then form the semidirect product group $G = H \rtimes V$, whose composition law is $(h, \mathbf{u})(k, \mathbf{v}) = (hk, k^{-1}\mathbf{u} + \mathbf{v})$. This G is the symmetry group we will consider in this section.

The group G has a natural left action on the joint outcome space $\mathcal{X} \times \mathcal{Y} = \mathbb{F}^2$, that is,

$$(h, \mathbf{u}) \begin{pmatrix} x \\ y \end{pmatrix} = h \begin{pmatrix} x + u_1 \\ y + u_2 \end{pmatrix}, \quad (60)$$

where the points of $\mathcal{X} \times \mathcal{Y} = \mathbb{F}^2$ have been written as columns. In this context, the joint outcome space $\mathcal{X} \times \mathcal{Y}$ is called the *finite phase-space* of the system, and the subgroup $V \subset G$ is the group of its translations $(d(1), \mathbf{u})$. The elements $(d(a), \mathbf{0}) \in H$ are diagonal symplectic transformations, while $(f(1), \mathbf{0})$ just reverses the components x and y changing sign to x (see e.g. [56] for more details on finite phase-spaces and their symmetries).

On the other hand, the group G has also a natural projective unitary representation on \mathcal{H} . In order to describe it, we first introduce the following unitary operators:

$$\begin{aligned} W(\mathbf{u})\phi(z) &= e^{\frac{2\pi i}{p} \text{tr } u_2(z-u_1)} \phi(z - u_1), \quad \forall \mathbf{u} \in \mathbb{F}^2, \\ D(a)\phi(z) &= \phi(a^{-1}z), \quad \forall a \in \mathbb{F}_* = \mathbb{F} \setminus \{0\}. \end{aligned}$$

The operators $W(\mathbf{u})$ constitute the *Weyl operators* associated with the phase-space translations, and $D(a)$ are the *squeezing operators* by the nonzero scalars. Collected together with the Fourier transform F , they satisfy the composition rules

$$\begin{aligned} W(\mathbf{u})W(\mathbf{v}) &= e^{\frac{2\pi i}{p} \text{tr } u_2 v_1} W(\mathbf{u} + \mathbf{v}), & D(a)D(b) &= D(ab), \\ F^2 &= D_{-1}, & FD(a)F^* &= D(a^{-1}), \\ D(a)W(\mathbf{u})D(a)^* &= W(d(a)\mathbf{u}), & FW(\mathbf{u})F^* &= e^{-\frac{2\pi i}{p} \text{tr } u_1 u_2} W(f(1)\mathbf{u}). \end{aligned}$$

Setting

$$U(d(a), \mathbf{u}) = D(a)W(\mathbf{u}), \quad U(f(a), \mathbf{u}) = D(a)FW(\mathbf{u}),$$

we obtain a projective unitary representation of G on \mathcal{H} . It is easily checked that

$$\begin{aligned} U(d(a), \mathbf{u})Q(x)U(d(a), \mathbf{u})^* &= Q(a(x + u_1)), \\ U(d(a), \mathbf{u})P(y)U(d(a), \mathbf{u})^* &= P(a^{-1}(y + u_2)), \\ U(f(a), \mathbf{u})Q(x)U(f(a), \mathbf{u})^* &= P(-a^{-1}(x + u_1)), \\ U(f(a), \mathbf{u})P(y)U(f(a), \mathbf{u})^* &= Q(a(y + u_2)). \end{aligned} \tag{61}$$

The action (60) satisfies the hypotheses (i.a) / (ii.a) of Theorem 8, with $S_G \equiv G$. Moreover, by (61) the two sharp observables $A \equiv Q$ and $B \equiv P$ satisfy the conditions (i.b) / (ii.b) of the same theorem. Therefore, by Corollary 9 we conclude that the set $\mathcal{M}_{\text{inc}}(Q, P) = \mathcal{M}_{\text{ed}}(Q, P)$ contains a G -covariant element M_0 .

Since in particular the bi-observable M_0 is covariant with respect to the group V of the phase-space translations, it must be of the form

$$M_\tau(x, y) = \frac{1}{d} W((x, y)^T) \tau W((x, y)^T)^*, \quad \forall x, y \in \mathbb{F}, \tag{62}$$

i.e. $M_0 = M_{\tau_0}$ for some state $\tau_0 \in \mathcal{S}(\mathcal{H})$ [37, Theor. 4.5.3]. We call an observable M_τ of the form (62) the *V-covariant phase-space observable generated by the state τ* . Since M_0 is also H -covariant and H is the stability subgroup of G at $(0, 0)$, we see that $\tau_0 = d M_0(0, 0)$ can be any state commuting with the restriction $U|_H$ of the representation U to H .

By [49, Propositions 1 and 2], the marginals of a V -covariant phase-space observable M_τ are

$$M_{\tau[1]}(x) = \sum_{z \in \mathbb{F}} Q^\tau(z - x)Q(z), \quad M_{\tau[2]}(y) = \sum_{z \in \mathbb{F}} P^\tau(z - y)P(z). \tag{63}$$

Now, the fact that τ_0 commutes with $U|_H$ and the covariance relations (61) imply

$$Q^{\tau_0}(x) = \text{Tr} [\tau_0 U(f(-1), \mathbf{0}) Q(x) U(f(-1), \mathbf{0})^*] = P^{\tau_0}(x), \quad \forall x \in \mathbb{F},$$

and

$$Q^{\tau_0}(x) = \text{Tr} [\tau_0 U(d(a), \mathbf{0}) Q(x) U(d(a), \mathbf{0})^*] = Q^{\tau_0}(ax), \quad \forall x \in \mathbb{F}, a \in \mathbb{F}_*.$$

By the second relation, the probability Q^{τ_0} is constant on the two subsets $\{0\}$ and \mathbb{F}_* of \mathbb{F} , which are the orbits of the action of the multiplicative group \mathbb{F}_* on \mathbb{F} . Therefore, we can write Q^{τ_0} as a linear combination of the two functions δ_0 and $u_{\mathbb{F}} - \delta_0/d$. The normalization of Q^{τ_0} requires

$$Q^{\tau_0} = \lambda_0 \delta_0 + (1 - \lambda_0) u_{\mathbb{F}}$$

for some real λ_0 . On the other hand, we must have $\lambda_0 \in [-1/(d - 1), 1]$ by the positivity constraint. Equations (63) with $\tau = \tau_0$ then give

$$M_{0[1]} = \lambda_0 Q + (1 - \lambda_0) U_{\mathbb{F}} \equiv Q_{\lambda_0}, \quad M_{0[2]} = \lambda_0 P + (1 - \lambda_0) U_{\mathbb{F}} \equiv P_{\lambda_0},$$

where $U_{\mathbb{F}}$ is the uniform observable on \mathbb{F} . If $\lambda_0 \geq 0$, then Q_{λ_0} and P_{λ_0} have the simple physical interpretation as uniformly noisy versions of Q and P with noise intensities $1 - \lambda_0$, as it was explained in Section 2.3

(see (14)). However, we can not exclude that λ_0 takes its value in the negative interval $[-1/(d-1), 0)$, where this interpretation does not apply.

A straightforward extension of the argument in [50, Prop. 5] from the cyclic field \mathbb{Z}_p to the finite field \mathbb{F} yields the following compatibility condition for two uniformly noisy versions of Q and P (see also Example 1 therein).

Proposition 15. *Suppose $\lambda \in [0, 1]$, and let $Q_\lambda = \lambda Q + (1 - \lambda)U_{\mathbb{F}}$ and $P_\lambda = \lambda P + (1 - \lambda)U_{\mathbb{F}}$ be two noisy versions of Q and P with the same noise intensity $1 - \lambda$. Then Q_λ and P_λ are compatible if and only if*

$$\lambda \leq \lambda_{\max} = \frac{d + \sqrt{d} - 2}{2(d-1)}. \quad (64)$$

When in the previous bound the equality is attained, $Q_{\lambda_{\max}}$ and $P_{\lambda_{\max}}$ have a unique joint measurement in the whole set $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$, which is the V -covariant phase-space observable $M_{\tau_{\max}}$ generated by the pure state

$$\tau_{\max} = \frac{\sqrt{d}}{2(1 + \sqrt{d})} |\delta_0 + F\delta_0\rangle \langle \delta_0 + F\delta_0|. \quad (65)$$

As a consequence, for the two marginals Q_{λ_0} and P_{λ_0} of the optimal approximate joint measurement M_0 , the inequalities $-1/(d-1) \leq \lambda_0 \leq \lambda_{\max}$ must hold. Note that the state τ_{\max} commutes with $U|_H$, hence it is a valid candidate for generating the G -covariant phase-space observable M_0 . Indeed, in the next theorem we show that actually $M_0 = M_{\tau_{\max}}$ and $\lambda_0 = \lambda_{\max}$, and we use this fact to obtain a lower bound for the entropic indexes $c_{\text{inc}}(Q, P) = c_{\text{ed}}(Q, P)$ (recall that Q and P are sharp).

Theorem 16. *For the two sharp observables Q and P defined in (56) and (57), we have*

$$\log \frac{2\sqrt{d}}{\sqrt{d} + 1} \leq c_{\text{inc}}(Q, P) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} [S(Q^\rho \| Q_{\lambda_{\max}}^\rho) + S(P^\rho \| P_{\lambda_{\max}}^\rho)] \leq 2 \log \frac{2(d+1)}{d+3}, \quad (66)$$

with λ_{\max} defined in (64). Moreover $M_{\tau_{\max}}$, defined by (62) and (65), is the unique G -covariant observable in $\mathcal{M}_{\text{inc}}(Q, P)$, and, if $p \neq 2$,

$$\mathcal{M}_{\text{inc}}(Q, P) = \{M_{\tau_{\max}}\}. \quad (67)$$

Proof. For a G -covariant optimal approximate joint measurement M_0 as in the previous discussion, we have

$$c_{\text{inc}}(Q, P) = D(Q, P | M_0) = \sup_{\rho} [S(Q^\rho \| Q_{\lambda_0}^\rho) + S(P^\rho \| P_{\lambda_0}^\rho)] =: D(\lambda_0),$$

where the map $\lambda \mapsto D(\lambda) = \sup_{\rho} [S(Q^\rho \| Q_{\lambda}^\rho) + S(P^\rho \| P_{\lambda}^\rho)]$ is a strictly decreasing function of $\lambda \in [-1/(d-1), 1]$ by Lemma 17 below. This fact and Proposition 15 then imply $\lambda_0 = \lambda_{\max}$. Moreover, the fact that $M_{\tau_{\max}}$ is the unique joint observable of $Q_{\lambda_{\max}}$ and $P_{\lambda_{\max}}$ imposes $M_0 = M_{\tau_{\max}}$. Therefore, $M_{\tau_{\max}}$ is the unique G -covariant observable in $\mathcal{M}_{\text{inc}}(Q, P)$, and

$$c_{\text{inc}}(Q, P) = D(\lambda_{\max}) = \sup_{\rho} [S(Q^\rho \| Q_{\lambda_{\max}}^\rho) + S(P^\rho \| P_{\lambda_{\max}}^\rho)].$$

The first inequality in (66) then follows evaluating the bound (68) of Lemma 17 at $\lambda = \lambda_{\max}$. On the other hand, the second inequality is the general bound for $c_{\text{inc}}(Q, P)$ given in (17). If M is any observable in the optimal set $\mathcal{M}_{\text{inc}}(Q, P)$, its covariant version M_G is still in $\mathcal{M}_{\text{inc}}(Q, P)$ by Corollary 9, hence $M_G = M_{\tau_{\max}}$ by the previous part. By (22), $M(x, y) \leq |G|M_G(x, y) = |G|M_{\tau_{\max}}(x, y)$ for all x, y . Since $M_{\tau_{\max}}(x, y)$ has rank 1, it must then be $M(x, y) = f(x, y)M_{\tau_{\max}}(x, y)$ for some function $f : \mathbb{F}^2 \rightarrow [0, |G|]$. The two normalization constraints $\sum_{x, y} M_{\tau_{\max}}(x, y) = \mathbb{1}$ and $\sum_{x, y} f(x, y)M_{\tau_{\max}}(x, y) = \sum_{x, y} M(x, y) = \mathbb{1}$ impose constraints on the coefficients $f(x, y)$. If $d = p^n$ is odd, these constraints are enough to imply that $f(x, y) = 1$ for all x, y . Indeed, this follows since in this case the observable $M_{\tau_{\max}}$ is informationally complete. Indeed, for $d = p$ odd, this is proved in [50, Prop. 9]. In the more general case $d = p^n$ odd, the same proof still holds, as it relies on the fact that the inverse Weyl transform of τ_{\max}

$$\hat{\tau}_{\max}(\mathbf{u}) := \text{Tr} \{ \tau_{\max} W(\mathbf{u}) \} = \frac{\sqrt{d}}{2(1 + \sqrt{d})} \left[\delta_0(u_1) + \delta_0(u_2) + \frac{1}{\sqrt{d}} \left(e^{-\frac{2\pi i}{p} \text{tr } u_1 u_2} + 1 \right) \right]$$

is nonzero for all $\mathbf{u} \in \mathbb{F}^2$ (see [58, Prop. 12]). The proof of the uniqueness statement is thus concluded. \square

Remark 8. 1. The two bounds in (66) are not asymptotically optimal for $d \rightarrow \infty$, as the lower bound tends to 1 while the upper bound goes to 2.

2. In Theorem 16, the uniqueness property of the optimal approximate joint measurement $M_{\tau_{\max}}$ in odd prime power dimensions should be compared with the measurement uncertainty region for two qudit observables found in [11, Sect. V.C]. In particular, we remark that there is a whole family of V -covariant phase-space observables M_ρ saturating the uncertainty bound of [11, Eq. 38]. Our optimal observable $M_{\tau_{\max}}$ just corresponds to one of them, that is, the case $\rho = \tau_{\max}$, which in [11, Eq. 38] gives the Wasserstein spreads $d(\rho^P) = d(\rho^Q) = (\sqrt{d} - 1)/(2\sqrt{d})$.

3. When $d = 2^n$ with $n \geq 2$, it is not clear whether or not there exist noncovariant observables in $\mathcal{M}_{\text{inc}}(\mathbf{Q}, \mathbf{P})$. However, in the simplest case $d = 2$ we have already shown that $\mathcal{M}_{\text{inc}}(\mathbf{Q}, \mathbf{P}) = \{M_{\tau_{\max}}\}$ with a different argument (Proposition 14).

In the proof of Theorem 16, we made use of the following lemma.

Lemma 17. *The function*

$$D : [-1/(d-1), 1] \rightarrow [0, +\infty] \quad D(\lambda) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} [S(\mathbf{Q}^\rho \| \mathbf{Q}_\lambda^\rho) + S(\mathbf{P}^\rho \| \mathbf{P}_\lambda^\rho)]$$

is convex and strictly decreasing, and it satisfies the inequality

$$D(\lambda) \geq \log \frac{d}{(d-1)\lambda + 1}, \quad \forall \lambda \in [0, 1]. \quad (68)$$

Proof. The two maps $\lambda \mapsto S(\mathbf{Q}^\rho \| \mathbf{Q}_\lambda^\rho)$ and $\lambda \mapsto S(\mathbf{P}^\rho \| \mathbf{P}_\lambda^\rho)$ are convex for $\lambda \in [-1/(d-1), 1]$ by item (i) of Proposition 2. Hence D is convex, being the supremum of the sum of two convex functions.

Since $\mathbf{Q}_1 = \mathbf{Q}$ and $\mathbf{P}_1 = \mathbf{P}$, we have $D(1) = 0$. Thus, for $\lambda' < \lambda < 1$, the convexity of D implies

$$D(\lambda) \leq \frac{1-\lambda}{1-\lambda'} D(\lambda') + \frac{\lambda-\lambda'}{1-\lambda'} D(1) = \frac{1-\lambda}{1-\lambda'} D(\lambda') < D(\lambda'),$$

which shows that D is strictly decreasing. Finally, evaluating the function inside the sup at any eigenstate $\rho = |\delta_x\rangle\langle\delta_x|$ of \mathbf{Q} ,

$$D(\lambda) \geq \log \frac{d}{(d-1)\lambda + 1} > 0, \quad \forall \lambda \in [-1/(d-1), 1).$$

□

Example 1 (Two orthogonal spin-1/2 components). Let us consider as reference observables the two sharp spin-1/2 components $\mathbf{X}, \mathbf{Y} \in \mathcal{M}(\{+1, -1\})$ associated with the first two Pauli matrices, defined in (37). This is the easiest example of sharp observables related to two Fourier conjugate MUB. To see this, take the cyclic field $\mathbb{F} = \mathbb{Z}_2 \equiv \{0, 1\}$, corresponding to the choice $d = p = 2$, $n = 1$, $\text{tr } x = x$, and identify the observables $\mathbf{Q}(x) = \mathbf{X}((-1)^x)$ and $\mathbf{P}(y) = \mathbf{Y}((-1)^y)$ ($x, y = 0, 1$) by setting

$$\sigma_1 = |\delta_0\rangle\langle\delta_0| - |\delta_1\rangle\langle\delta_1|, \quad \sigma_2 = |\delta_0\rangle\langle\delta_1| + |\delta_1\rangle\langle\delta_0|.$$

With this identification, the discrete Fourier transform becomes

$$F = \frac{\sigma_1 + \sigma_2}{\sqrt{2}} \equiv \text{ie}^{-i\pi \mathbf{n} \cdot \boldsymbol{\sigma} / 2}.$$

We already found in (47) the optimal joint observable of \mathbf{X} and \mathbf{Y} , together with the value of the entropic incompatibility degree. Note that these are precisely the bi-observable and the lower bound found in Theorem 16.

4 Entropic divergence and incompatibility degree for n observables

Both the entropic coefficients (7) and (9) can be generalized to the case of more than two observables. However, in the case of $c_{\text{ed}}(A_1, \dots, A_n)$ an order of observation has to be fixed, and one needs to point out the subset of the observables for which imprecise measurements are allowed (the analogues of the observable A in the binary case of $c_{\text{ed}}(A, B)$) from those observables that are kept fixed and get disturbed by the other measurements (similar to B in $c_{\text{ed}}(A, B)$). Thus, different definitions of c_{ed} are possible in the n -observable case. This leads us to generalize only the entropic incompatibility degree $c_{\text{inc}}(A_1, \dots, A_n)$, whose definition is straightforward and gives a lower bound for c_{ed} , independently of its definition.

Let A_1, \dots, A_n be n fixed observables with outcome sets $\mathcal{X}_1, \dots, \mathcal{X}_n$, respectively. As usual, we assume that all the sets \mathcal{X}_i are finite. The observables with outcomes in the product set $\mathcal{X}_1 \times \dots \times \mathcal{X}_n =: \mathcal{X}_{1\dots n}$ are called multi-observables, and we use the notation $\mathcal{M}(\mathcal{X}_{1\dots n})$ for the set of all such observables. If $M \in \mathcal{M}(\mathcal{X}_{1\dots n})$, its i -th marginal observable is the element $M_{[i]} \in \mathcal{M}(\mathcal{X}_i)$, with

$$M_{[i]}(x) = \sum_{x_j \in \mathcal{X}_j: j \neq i} M(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n).$$

The notion of compatibility straightforwardly extends to the case of n observables. Moreover, considering $M \in \mathcal{M}(\mathcal{X}_{1\dots n})$ as an *approximate joint measurement* of A_1, \dots, A_n , for any $\rho \in \mathcal{S}(\mathcal{H})$ the total amount of lost information in the distribution approximations $A_i^\rho \simeq M_{[i]}^\rho$ is the sum

$$S[A_1, \dots, A_n \| M](\rho) = \sum_{i=1}^n S(A_i^\rho \| M_{[i]}^\rho). \quad (69)$$

We have the following generalization of Definitions 1 and 2.

Definition 4. The *entropic divergence* of $M \in \mathcal{M}(\mathcal{X}_{1\dots n})$ from (A_1, \dots, A_n) is

$$D(A_1, \dots, A_n \| M) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} \sum_{i=1}^n S(A_i^\rho \| M_{[i]}^\rho). \quad (70)$$

Definition 5. The *entropic incompatibility degree* of the observables A_1, \dots, A_n is

$$c_{\text{inc}}(A_1, \dots, A_n) = \inf_{M \in \mathcal{M}(\mathcal{X}_{1\dots n})} D(A_1, \dots, A_n \| M). \quad (71)$$

We still denote by

$$\mathcal{M}_{\text{inc}}(A_1, \dots, A_n) = \arg \min_{M \in \mathcal{M}(\mathcal{X}_{1\dots n})} D(A_1, \dots, A_n \| M)$$

the corresponding set of minimizing multi-observables, and refer to its elements as the *optimal approximate joint measurements* of A_1, \dots, A_n . As in the case with $n = 2$, the optimality of a multi-observable M depends only on its marginals $M_{[i]}$, since the entropic divergence itself depends only on such marginals. We have the following extension of Theorems 3, 6 and 7.

Theorem 18. Let $A_i \in \mathcal{M}(\mathcal{X}_i)$, $i = 1, \dots, n$, be the reference observables. The entropic divergence and incompatibility degree satisfy the following properties.

- (i) The function $S[A_1, \dots, A_n \| M] : \mathcal{S}(\mathcal{H}) \rightarrow [0, +\infty]$ is convex and LSC, $\forall M \in \mathcal{M}(\mathcal{X}_{1\dots n})$.
- (ii) The function $D(A_1, \dots, A_n \| \cdot) : \mathcal{M}(\mathcal{X}_{1\dots n}) \rightarrow [0, +\infty]$ is convex and LSC.
- (iii) For any $M \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, the following three statements are equivalent:
 - (a) $D(A_1, \dots, A_n \| M) < +\infty$,
 - (b) $\ker M_{[i]}(x) \subseteq \ker A_i(x) \quad \forall x, \forall i$,

(c) $S[A_1, \dots, A_n \| M]$ is bounded and continuous.

(iv) $D(A_1, \dots, A_n \| M) = \max_{\substack{\rho \in \mathcal{S}(\mathcal{H}) \\ \rho \text{ pure}}} \sum_{i=1}^n S(A_i^\rho \| M_{[i]}^\rho)$, where the maximum can be any value in $[0, +\infty]$.

(v) The quantities $D(A_1, \dots, A_n \| M)$ and $c_{\text{inc}}(A_1, \dots, A_n)$ are invariant under an overall unitary conjugation of the observables A_1, \dots, A_n and M , and they do not depend on the labelling of the outcomes in $\mathcal{X}_1, \dots, \mathcal{X}_n$.

(vi) $c_{\text{inc}}(A_{\sigma(1)}, \dots, A_{\sigma(n)}) = c_{\text{inc}}(A_1, \dots, A_n)$ for any permutation σ of the index set $\{1, \dots, n\}$.

(vii) The entropic incompatibility coefficient $c_{\text{inc}}(A_1, \dots, A_n)$ is always finite, and it satisfies

$$c_{\text{inc}}(A_1, \dots, A_n) \leq \sum_{i=1}^n \log |\mathcal{X}_i| - \inf_{\rho \in \mathcal{S}(\mathcal{H})} \sum_{i=1}^n H(A_i^\rho), \quad (72)$$

$$c_{\text{inc}}(A_1, \dots, A_n) \leq \sum_{i=1}^n \log \frac{n(d+1)}{d+n+d(n-1) \min_{x \in \mathcal{X}_i} A_i^{\rho_0}(x)} \leq n \log n. \quad (73)$$

(viii) The set $\mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$ is a nonempty convex compact subset of $\mathcal{M}(\mathcal{X}_1 \dots \mathcal{X}_n)$.

(ix) $c_{\text{inc}}(A_1, \dots, A_n) = 0$ if and only if the observables A_1, \dots, A_n are compatible, and in this case $\mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$ is the set of all the joint measurements of A_1, \dots, A_n .

(x) If $A_{n+1} \in \mathcal{M}(\mathcal{X}_{n+1})$ is another observable, then $c_{\text{inc}}(A_1, \dots, A_{n+1}) \geq c_{\text{inc}}(A_1, \dots, A_n)$.

Proof. The proofs of properties (i)–(vi), (viii) and (ix) are straightforward extensions of the analogous ones for two observables.

The first upper bound in property (vii) follows by evaluating the entropic divergence of the uniform observable $U = (u_{\mathcal{X}_1} \otimes \dots \otimes u_{\mathcal{X}_n}) \mathbb{1}$ from (A_1, \dots, A_n)

$$\begin{aligned} c_{\text{inc}}(A_1, \dots, A_n) &\leq D(A_1, \dots, A_n \| U) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} \sum_{i=1}^n S(A_i^\rho \| u_{\mathcal{X}_i}) \\ &= \sup_{\rho \in \mathcal{S}(\mathcal{H})} \sum_{i=1}^n [\log |\mathcal{X}_i| - H(A_i^\rho)] \quad \text{by item (iii) of Proposition 1} \end{aligned}$$

which then yields (72).

The second upper bound in property (vii) follows by using an approximate cloning argument, just as in the case of only two observables. Indeed, the optimal approximate n -cloning channel is the map

$$\Phi : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H}^{\otimes n}), \quad \Phi(\rho) = \frac{d!n!}{(d+n-1)!} S_n(\rho \otimes \mathbb{1}^{\otimes(n-1)}) S_n,$$

where S_n is the orthogonal projection of $\mathcal{H}^{\otimes n}$ onto its symmetric subspace $\text{Sym}(\mathcal{H}^{\otimes n})$ [46]. Evaluating the marginals of the multi-observable $M_{\text{cl}} = \Phi^*(A_1 \otimes \dots \otimes A_n)$, we obtain the noisy versions

$$M_{\text{cl}}[i] = A_i \lambda_{\text{cl}}, \quad \text{where} \quad \lambda_{\text{cl}} = \frac{d+n}{n(d+1)}$$

(see [47]). With a computation similar to the one for obtaining the bound (16) in Section 2.3, we get

$$c_{\text{inc}}(A_1, \dots, A_n) \leq D(A_1, \dots, A_n \| M_{\text{cl}}) \leq \sum_{i=1}^n \log \frac{n(d+1)}{d+n+d(n-1) \min_{x \in \mathcal{X}_i} A_i^{\rho_0}(x)} \leq n \log n.$$

Finally, in order to prove property (x), take any $M' \in \mathcal{M}(\mathcal{X}_1 \times \dots \times \mathcal{X}_{n+1})$, and let

$$M(x_1, \dots, x_n) = \sum_{x \in \mathcal{X}_{n+1}} M'(x_1, \dots, x_n, x).$$

We have $M'_{[i]} = M_{[i]}$ for all $i = 1, \dots, n$, hence

$$\begin{aligned} c_{\text{inc}}(A_1, \dots, A_n) &\leq D(A_1, \dots, A_n \| M) = \sup_{\rho} \sum_{i=1}^n S(A_i^{\rho} \| M_{[i]}^{\rho}) \leq \sup_{\rho} \sum_{i=1}^{n+1} S(A_i^{\rho} \| M'_{[i]}^{\rho}) \\ &= D(A_1, \dots, A_{n+1} \| M'). \end{aligned}$$

Item (x) then follows by taking the infimum over M' . \square

Remark 9. If all the observables A_1, \dots, A_n are sharp, with $|\mathcal{X}_i| = d$ and $\text{rank } A_i(x) = 1$ for all $x \in \mathcal{X}_i$ and $i = 1, \dots, n$, the first inequality in (73) can be refined to

$$c_{\text{inc}}(A_1, \dots, A_n) \leq n \log \frac{n(d+1)}{d+2n-1}, \quad (74)$$

which is always better than the bound $c_{\text{inc}}(A_1, \dots, A_n) \leq n \log d$ given by (72).

Remark 10. The monotonicity property (x), which is specific of the many observable case, is another desirable feature for an incompatibility coefficient: the amount of incompatibility cannot decrease when an extra observable is added.

Finally, suppose the product space $\mathcal{X}_{1\dots n}$ carries the action of a finite symmetry group G , which also acts on the quantum system Hilbert space \mathcal{H} by means of a projective unitary representation U . These actions then extend to the set of states $\mathcal{S}(\mathcal{H})$, the set of probabilities $\mathcal{P}(\mathcal{X}_{1\dots n})$ and the set of multi-observables $\mathcal{M}(\mathcal{X}_{1\dots n})$ exactly as in Section 3.1. Similarly, for any $M \in \mathcal{M}(\mathcal{X}_{1\dots n})$, we can define its covariant version M_G . The content of Theorem 8 and Corollary 9 then can be translated to the case of n observables as follows.

Theorem 19. *Let $A_i \in \mathcal{M}(\mathcal{X}_i)$, $i = 1, \dots, n$, be the reference observables. Suppose the finite group G acts on both the output space $\mathcal{X}_{1\dots n}$ and the index set $\{1, \dots, n\}$, and it also acts with a projective unitary representation U on \mathcal{H} . Moreover, assume that G is generated by a subset $S_G \subseteq G$ such that, for every $g \in S_G$ and $i \in \{1, \dots, n\}$, there exists a bijective map $f_{g,i} : \mathcal{X}_i \rightarrow \mathcal{X}_{gi}$ for which*

$$(a) \quad g(x_1, \dots, x_n)_{gi} = f_{g,i}(x_i) \text{ for all } (x_1, \dots, x_n) \in \mathcal{X}_{1\dots n}$$

$$(b) \quad U_g A_i(x_i) U_g^* = A_{gi}(f_{g,i}(x_i)) \text{ for all } x_i \in \mathcal{X}_i.$$

Then

- $D(A_1, \dots, A_n \| gM) = D(A_1, \dots, A_n \| M)$ for all $M \in \mathcal{M}(\mathcal{X}_{1\dots n})$ and $g \in G$;
- the set $\mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$ is G -invariant;
- for any $M \in \mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$, we have $M_G \in \mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$;
- there exists a G -covariant observable in $\mathcal{M}_{\text{inc}}(A_1, \dots, A_n)$.

Proof. As in the proof of Theorem 8, it is not restrictive to assume that $S_G = G$. For all $p \in \mathcal{P}(\mathcal{X}_{1\dots n})$, condition (a) implies

$$\begin{aligned} gp_{[i]}(x_i) &= \sum_{x_j \in \mathcal{X}_j : j \neq i} gp(x_1, \dots, x_n) = \sum_{x_{gj} \in \mathcal{X}_{gj} : gj \neq i} p(f_{g^{-1},g1}(x_{g1}), \dots, f_{g^{-1},gn}(x_{gn})) \\ &= \sum_{y_j \in \mathcal{X}_j : gj \neq i} p(y_1, \dots, y_n) \quad \text{where } y_j = f_{g^{-1},gj}(x_{gj}) \\ &= \sum_{y_j \in \mathcal{X}_j : j \neq g^{-1}i} p(y_1, \dots, y_n) = p_{[g^{-1}i]}(y_{g^{-1}i}) \\ &= p_{[g^{-1}i]}(f_{g^{-1},i}(x_i)), \end{aligned}$$

and hence

$$\begin{aligned} (gp_{[1]} \otimes \cdots \otimes gp_{[n]})(x_1, \dots, x_n) &= \prod_{i=1}^n gp_{[i]}(x_i) = \prod_{i=1}^n p_{[g^{-1}i]}(f_{g^{-1},i}(x_i)) = \prod_{i=1}^n p_{[i]}(f_{g^{-1},gi}(x_{gi})) \\ &= g(p_{[1]} \otimes \cdots \otimes p_{[n]})(x_1, \dots, x_n). \end{aligned}$$

Therefore,

$$gp_{[1]} \otimes \cdots \otimes gp_{[n]} = g(p_{[1]} \otimes \cdots \otimes p_{[n]}). \quad (75)$$

On the other hand, by condition (b) we have $A_i^{g\rho}(x_i) = A_{g^{-1}i}^\rho(f_{g^{-1},i}(x_i))$, and then

$$\begin{aligned} (A_1^{g\rho} \otimes \cdots \otimes A_n^{g\rho})(x_1, \dots, x_n) &= \prod_{i=1}^n A_{g^{-1}i}^\rho(f_{g^{-1},i}(x_i)) = \prod_{i=1}^n A_i^\rho(f_{g^{-1},gi}(x_{gi})) \\ &= g(A_1^\rho \otimes \cdots \otimes A_n^\rho)(x_1, \dots, x_n), \end{aligned}$$

that is,

$$A_1^{g\rho} \otimes \cdots \otimes A_n^{g\rho} = g(A_1^\rho \otimes \cdots \otimes A_n^\rho). \quad (76)$$

Having established (75) and (76), the proof that $D(A_1, \dots, A_n \| gM) = D(A_1, \dots, A_n \| M)$ follows along the same lines of the proof of Theorem 8. The remaining statements are then proved as in Corollary 9. \square

In the next example of $n = 3$ orthogonal spin-1/2 components, we will use Theorem 19 in order to evaluate their entropic incompatibility degree and the corresponding covariant optimal approximate joint measurements. Three sharp orthogonal spin-1/2 components are the basic example of a maximal set of $d + 1$ MUB in a d -dimensional Hilbert space. It is then natural to ask whether similar arguments lead to find the incompatibility index $c_{\text{inc}}(Q_1, \dots, Q_{d+1})$ of a maximal set of $d + 1$ MUB Q_1, \dots, Q_{d+1} whenever such a set of MUB is known to exist, that is, for all prime powers d . However, very few is known in this general case; for example, it is not even known the minimal amount of noise $1 - \lambda$ which makes compatible all the noisy versions $Q_{1\lambda}, \dots, Q_{d+1\lambda}$.

Example 2 (Three orthogonal spin-1/2 components). Let the reference observables A_1, A_2, A_3 be three mutually orthogonal spin-1/2 components, that is, the sharp observables $X, Y, Z \in \mathcal{M}(\{+1, -1\})$ associated with the three Pauli matrices; the observables X, Y are given in (37), and

$$Z(z) = \frac{1}{2} (\mathbb{1} + z\sigma_3), \quad \forall z \in \mathcal{Z} := \{+1, -1\}. \quad (77)$$

Let $O \subset SO(3)$ be the order 24 octahedron group generated by the 90° rotations around the three axes: $SO = \{R_i(\pi/2), R_j(\pi/2), R_k(\pi/2)\}$. Note that for the dihedral groups introduced in Sections 3.2 we have $D_2 \subset D_4 \subset O$. Let us denote the three generators of O by $g_1 = R_i(\pi/2)$, $g_2 = R_j(\pi/2)$, $g_3 = R_k(\pi/2)$. By using again the spin-1/2 projective representation of $SO(3)$, which we now restrict to O , we have the relations

$$\begin{aligned} U_{g_1} X(x) U_{g_1}^* &= X(x), & U_{g_1} Y(y) U_{g_1}^* &= Z(y), & U_{g_1} Z(z) U_{g_1}^* &= Y(-z), \\ U_{g_2} X(x) U_{g_2}^* &= Z(-x), & U_{g_2} Y(y) U_{g_2}^* &= Y(y), & U_{g_2} Z(z) U_{g_2}^* &= X(z), \\ U_{g_3} X(x) U_{g_3}^* &= Y(x), & U_{g_3} Y(y) U_{g_3}^* &= X(-y), & U_{g_3} Z(z) U_{g_3}^* &= Z(z). \end{aligned} \quad (78)$$

Moreover, the natural action of O on the outcome space $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} = \{+1, -1\}^3$ is given by

$$g_1(x, y, z) = (x, -z, y), \quad g_2(x, y, z) = (z, y, -x), \quad g_3(x, y, z) = (-y, x, z),$$

and the action on the index set is

$$g_i i = i, \quad g_1 2 = 3, \quad g_1 3 = 2, \quad g_2 1 = 3, \quad g_2 3 = 1, \quad g_3 1 = 2, \quad g_3 2 = 1.$$

Then, the hypotheses of Theorem 19 are satisfied by setting

$$f_{g_1,1}(x) = x, \quad f_{g_1,2}(y) = y, \quad f_{g_1,3}(z) = -z, \quad f_{g_2,1}(x) = -x,$$

$$f_{g_2,2}(y) = y, \quad f_{g_2,3}(z) = z, \quad f_{g_3,1}(x) = x, \quad f_{g_3,2}(y) = -y, \quad f_{g_3,3}(z) = z.$$

By similar arguments as in the proofs of Proposition 11 and Corollary 12, one can prove that the most general O -covariant tri-observable in $\mathcal{M}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ has the form

$$M(x, y, z) = \frac{1}{8} [\mathbb{1} + c(x\sigma_1 + y\sigma_2 + z\sigma_3)] \quad \text{with} \quad |c| \leq \frac{1}{\sqrt{3}}. \quad (79)$$

A proof analogue to the one of Theorem 16 shows that there is a unique O -covariant optimal approximate joint measurement $M_0 \in \mathcal{M}_{\text{inc}}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$; this M_0 is given by (79) with $c = 1/\sqrt{3}$. Its marginals are

$$M_{0[1]}(x) = \frac{1}{2} \left[\mathbb{1} + \frac{x}{\sqrt{3}} \sigma_1 \right], \quad M_{0[2]}(y) = \frac{1}{2} \left[\mathbb{1} + \frac{y}{\sqrt{3}} \sigma_2 \right], \quad M_{0[3]}(z) = \frac{1}{2} \left[\mathbb{1} + \frac{z}{\sqrt{3}} \sigma_3 \right].$$

In Appendix B.3 we prove that

$$c_{\text{inc}}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = S(\mathcal{X}^{\rho_e} \| M_{0[1]}^{\rho_e}) + S(\mathcal{Y}^{\rho_e} \| M_{0[2]}^{\rho_e}) + S(\mathcal{Z}^{\rho_e} \| M_{0[3]}^{\rho_e})$$

with $\rho_e = \frac{1}{2} (\mathbb{1} \pm \sigma_i)$, $i = 1, 2, 3$. Then, we get

$$c_{\text{inc}}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \log \frac{2}{1 + 1/\sqrt{3}} = \log(3 - \sqrt{3}) \simeq 0.342497.$$

Remark 11. It should be remarked that, differently from the case with only $n = 2$ spins, the covariant optimal approximate joint measurement M_0 for $n = 3$ orthogonal spin-1/2 components is not unique in the set $\mathcal{M}_{\text{inc}}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$. Indeed, one can find a noncovariant tri-observable \tilde{M}_0 with the same marginals as M_0 , so that $D(\mathcal{X}, \mathcal{Y}, \mathcal{Z} \| \tilde{M}_0) = D(\mathcal{X}, \mathcal{Y}, \mathcal{Z} \| M_0)$, hence $\tilde{M}_0 \in \mathcal{M}_{\text{inc}}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, but $\tilde{M}_0 \neq M_0$. The following example is provided in [60, Sect. VI]

$$\begin{aligned} \tilde{M}_0(1, 1, -1) &= 2M_0(1, 1, -1), & \tilde{M}_0(1, -1, 1) &= 2M_0(1, -1, 1), \\ \tilde{M}_0(-1, 1, 1) &= 2M_0(-1, 1, 1), & \tilde{M}_0(-1, -1, -1) &= 2M_0(-1, -1, -1), \\ \tilde{M}_0(x, y, z) &= 0 \text{ otherwise.} \end{aligned}$$

A Examples of compatible but not sequentially compatible observables

First example from [23]. Apart from an exchange of A with B and some explicit computations, this example is taken from [23, Sect. III.C, and the end of Sect. III.A]. With $\mathcal{H} = \mathbb{C}^3$, $\mathcal{X} = \{1, 2\}$ and $\mathcal{Y} = \{1, \dots, 5\}$, the two reference observables are defined by

$$\begin{aligned} A(1) &= \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & A(2) &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \\ B(1) &= \frac{1}{4} \begin{pmatrix} 2 & 0 & -\sqrt{2} \\ 0 & 0 & 0 \\ -\sqrt{2} & 0 & 1 \end{pmatrix}, & B(2) &= \frac{1}{10} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 4 \end{pmatrix}, & B(3) &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ B(4) &= \frac{1}{10} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix}, & B(5) &= \frac{1}{4} \begin{pmatrix} 2 & 0 & \sqrt{2} \\ 0 & 0 & 0 \\ \sqrt{2} & 0 & 1 \end{pmatrix}. \end{aligned}$$

These two observables are compatible and a joint observable is

$$\begin{aligned} M(1, 1) &= B(1), \quad M(1, 5) = B(5), \quad M(2, 2) = B(2), \quad M(2, 3) = B(3), \quad M(2, 4) = B(4), \\ M(1, 2) &= M(1, 3) = M(1, 4) = M(2, 1) = M(2, 5) = 0. \end{aligned}$$

One can check that $M_{[1]} = A$ and $M_{[2]} = B$, which implies $c_{\text{inc}}(A, B) = 0$. Moreover, in [23, Sect. III.C] it is proved that: (1) there exists an instrument implementing B which does not disturb A; (2) any instrument implementing A disturbs B. By item (vi) of Theorem 6, this implies $c_{\text{ed}}(B, A) = 0$ and $c_{\text{ed}}(A, B) > 0$.

Second example from [23]. This is the first example of [23, Sect. III.A], which we report in the particular case in which the noise parameters are fixed and equal: $\mu = \nu =: \lambda$, $\lambda \in (\frac{1}{2}, \frac{2}{3}]$. The observables are two-valued ($\mathcal{X} = \mathcal{Y} = \{1, 2\}$), and are built up by using two noncommuting orthogonal projections P and Q : $[P, Q] \neq 0$. The joint observable M and its marginals are given by

$$M(1, 1) = (1 - \lambda)\mathbb{1}, \quad M(1, 2) = (2\lambda - 1)P, \quad M(2, 1) = (2\lambda - 1)Q,$$

$$M(2, 2) = \left(1 - \frac{3}{2}\lambda\right)(P + Q) + \frac{\lambda}{2}(\mathbb{1} - P + \mathbb{1} - Q),$$

$$A(1) = M_{[1]}(1) = \lambda P + (1 - \lambda)(\mathbb{1} - P), \quad A(2) = M_{[1]}(2) = \lambda(\mathbb{1} - P) + (1 - \lambda)P,$$

$$B(1) = M_{[2]}(1) = \lambda Q + (1 - \lambda)(\mathbb{1} - Q), \quad B(2) = M_{[2]}(2) = \lambda(\mathbb{1} - Q) + (1 - \lambda)Q.$$

The observables A and B are compatible by construction, and so $c_{\text{inc}}(A, B) = 0$. In [23], it is proved that there does not exist an instrument implementing A which does not disturb B ; it follows that $c_{\text{ed}}(A, B) > 0$ and, by exchanging P and Q , also $c_{\text{ed}}(B, A) > 0$.

B Supplementary results for spin-1/2 observables

In this appendix, the reference observables A and B are the two spin components defined in (26), (27). Moreover, as in (37), (77), we use the symbols X, Y, Z for the sharp observables associated with the orthogonal spin components along the three coordinate axes.

B.1 Evaluating c_{inc} for two orthogonal components

We parameterize the states by the points in the Bloch sphere:

$$\rho = \frac{1}{2}(\mathbb{1} + \mathbf{v} \cdot \boldsymbol{\sigma}), \quad \mathbf{v} = \text{Tr} \{\boldsymbol{\sigma} \rho\}, \quad |\mathbf{v}| \leq 1.$$

Then, we get

$$\begin{aligned} X^\rho(x) &= \frac{1}{2}(1 + xv_1), & Y^\rho(y) &= \frac{1}{2}(1 + yv_2), \\ M_{0[1]}^\rho(x) &= \frac{1}{2}\left(1 + \frac{x}{\sqrt{2}}v_1\right), & M_{0[2]}^\rho(y) &= \frac{1}{2}\left(1 + \frac{y}{\sqrt{2}}v_2\right), \end{aligned}$$

where M_0 is the optimal D_4 -covariant observable defined in (47). By setting

$$s(v) = (1 + v) \ln \frac{1 + v}{1 + v/\sqrt{2}} + (1 - v) \ln \frac{1 - v}{1 - v/\sqrt{2}}, \quad (80)$$

we have

$$(2 \ln 2) S(X^\rho \| M_{0[i]}^\rho) = s(v_i), \quad (2 \ln 2) S(X^\rho \otimes Y^\rho \| M_{0[1]}^\rho \otimes M_{0[2]}^\rho) = s(v_1) + s(v_2).$$

As $s(v)$ is a continuous even function of v , we can study it only for $v \in [0, 1]$. Moreover, $s(v)$ is proportional to a relative entropy for a couple of binary probabilities; then, it is convex and takes its minimum when the probabilities are equal, i.e. for $v = 0$, which gives $s(0) = 0$; furthermore, $s(v)$ increases with increasing difference between the two probabilities and the maximum is in $v = 1$, where $s(1) = 2 \ln[2(2 - \sqrt{2})]$. Then, the maximum of $s(v_1) + s(v_2)$ has to be searched for $v_3 = 0$ and $v_1^2 + v_2^2 = 1$; accordingly, we set $v_1 = \cos \phi$ and $v_2 = \sin \phi$, $\phi \in [0, \pi/2]$.

The best way to maximize $s(v_1) + s(v_2)$ is by means of a suitable integral representation. Namely, by taking the derivative of $(1 + v) \ln[(1 + v)/(1 + \lambda v)] + (1 - v) \ln[(1 - v)/(1 - \lambda v)]$ with respect to the new variable λ , and integrating the result from $\lambda = 1/\sqrt{2}$ to $\lambda = 1$, we obtain

$$s(v) = \int_{1/\sqrt{2}}^1 \frac{2v^2(1 - \lambda)}{1 - \lambda^2 v^2} d\lambda. \quad (81)$$

Then, by differentiation and simple computations, we get

$$f(\phi) := \frac{d}{d\phi}(s(\cos \phi) + s(\sin \phi)) = -\sin(4\phi) \int_{1/\sqrt{2}}^1 \frac{\lambda^2(1-\lambda)(2-\lambda^2)}{(1-\lambda^2(\sin \phi)^2)^2(1-\lambda^2(\cos \phi)^2)^2} d\lambda.$$

The integrand is nonnegative for all $\lambda \in [1/\sqrt{2}, 1]$ and $\phi \in [0, \pi/2]$. We then see that $f(\phi) < 0$ for $0 < \phi < \pi/4$, $f(\pi/4) = 0$, $f(\phi) > 0$ for $\pi/4 < \phi < \pi/2$. So, the point $\phi = \pi/4$ gives a minimum of $s(\cos \phi) + s(\sin \phi)$, while we have two equal maxima in $\phi = 0$ and $\phi = \pi/2$. This means that $s(v_1) + s(v_2)$ has its maxima in $v_1 = \pm 1$ with $v_2 = v_3 = 0$, and in $v_2 = \pm 1$ with $v_1 = v_3 = 0$; these points correspond to the eigenstates of σ_1 or σ_2 . Then, by direct computation we get the value (48). The point $v_1 = v_2 = 1/\sqrt{2}$ corresponds to a minimum among all the pure states with $v_3 = 0$.

B.2 A lower bound for the incompatibility degree

To compute the lower bound (50), we have to minimize over $c_2 = \gamma/\sqrt{2}$ the quantity

$$S[A, B \| M_\gamma](\rho_e) = \log \frac{2}{1 + a_2 c_2 + a_1/\sqrt{2}} + \frac{1}{2} (1 + 2a_1 a_2) \log \frac{1 + 2a_1 a_2}{1 + a_1 c_2 + a_2/\sqrt{2}} + \frac{1}{2} (1 - 2a_1 a_2) \log \frac{1 - 2a_1 a_2}{1 - a_1 c_2 - a_2/\sqrt{2}}. \quad (82)$$

By setting $\ell := a_1 c_2 + a_2/\sqrt{2}$ and $f(\ell) := (\ln 2) S[A, B \| M_\gamma](\rho_e)$, we get

$$f(\ell) = \ln \frac{2\sqrt{2} a_1}{\sqrt{2} a_1 + \sqrt{2} a_2 \ell + a_1^2 - a_2^2} + \frac{1}{2} (1 + 2a_1 a_2) \ln \frac{1 + 2a_1 a_2}{1 + \ell} + \frac{1}{2} (1 - 2a_1 a_2) \ln \frac{1 - 2a_1 a_2}{1 - \ell}, \quad (83)$$

whose derivative is

$$f'(\ell) = -\frac{\sqrt{2} a_2}{\sqrt{2} a_1 + \sqrt{2} a_2 \ell + a_1^2 - a_2^2} + \frac{\ell - 2a_1 a_2}{1 - \ell^2}.$$

Remark 12. For $\alpha \rightarrow 0$ one gets $a_1 = a_2 = 1/\sqrt{2}$, the two observable become compatible and, indeed, the minimum is zero in the point $\ell = \sqrt{2} a_2$, i.e. $c_2 = a_2$.

Remark 13. For $\alpha = \pi/2$, i.e. $a_2 = 0$ and $a_1 = 0$, we immediately get that the expression (83) has a unique minimum in $\ell = 0$, which gives $c_2 = 0$ and the value (48) for the incompatibility degree.

For $\alpha \neq \pi/2$, the zeros of the derivative satisfy the algebraic equation

$$\ell^2 + \frac{u}{\sqrt{2} a_2} \ell - 1 - u = 0,$$

where u is defined in (53). By solving the algebraic equation and studying the sign of the derivative, we find that the minimum of (83) is at the point (52) and that the corresponding value of c_2 is (51). By using this result and $2a_1 a_2 = \cos \alpha$, we get the lower bound (54).

B.3 Evaluating c_{inc} for three orthogonal components

For the O -covariant optimal approximate joint measurement $M_0 \in \mathcal{M}_{\text{inc}}(X, Y, Z)$ found in Example 2, along the same lines of Section B.1, we get

$$(2 \ln 2) S(X^\rho \otimes Y^\rho \otimes Z^\rho \| M_{0[1]}^\rho \otimes M_{0[2]}^\rho \otimes M_{0[3]}^\rho) = s(v_1) + s(v_2) + s(v_3).$$

where $s(v)$ is now the function

$$s(v) = (1 + v) \ln \frac{1 + v}{1 + v/\sqrt{3}} + (1 - v) \ln \frac{1 - v}{1 - v/\sqrt{3}}. \quad (84)$$

As before, the maximum of $s(v_1) + s(v_2) + s(v_3)$ has to be searched for $v_1^2 + v_2^2 + v_3^2 = 1$ and we can restrict $v_i \in [0, 1]$. For this reason, we set $v_1 = \cos \phi \sin \theta$, $v_2 = \sin \phi \sin \theta$, $v_3 = \cos \theta$, with $\phi, \theta \in [0, \pi/2]$.

By using the integral representation (81) with λ integrated from $1/\sqrt{3}$ to 1, we get

$$\begin{aligned} \frac{\partial}{\partial \phi} (s(\cos \phi \sin \theta) + s(\sin \phi \sin \theta) + s(\cos \theta)) \\ = -\sin(4\phi)(\sin \theta)^4 \int_{1/\sqrt{3}}^1 \frac{\lambda^2(1-\lambda)(2-\lambda^2)}{(1-\lambda^2 v_1^2)^2 (1-\lambda^2 v_2^2)^2} d\lambda; \end{aligned}$$

similar computations give the derivative with respect to θ . By the same arguments as in the case of two components, we obtain that the maximum of $s(\cos \phi \sin \theta) + s(\sin \phi \sin \theta) + s(\cos \theta)$ is for $\phi = 0, \pi/2$ and $\theta = 0, \pi/2$.

References

- [1] Ozawa, M.: *Position measuring interactions and the Heisenberg uncertainty principle*, Phys. Lett. A **299** (2002) 1–7.
- [2] Ozawa, M.: *Physical content of Heisenberg’s uncertainty relation: limitation and reformulation*, Phys. Lett. A **318** (2003) 21–29.
- [3] Ozawa, M.: *Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement*, Phys. Rev. A **67** (2003) 042105.
- [4] Ozawa, M.: *Uncertainty relations for joint measurements of noncommuting observables*, Phys. Lett. A **320** (2004) 367–374.
- [5] Werner, R.F.: *The uncertainty relation for joint measurement of position and momentum*, Quantum Inf. Comput. **4** (2004) 546–562.
- [6] Busch, P., Lahti, P., Werner, R.F.: *Measurement uncertainty relations*, J. Math. Phys. **55** (2014) 042111.
- [7] Busch, P., Lahti, P., Werner, R.F.: *Colloquium: Quantum root-mean-square error and measurement uncertainty relations*, Rev. Mod. Phys. **86** (2014) 1261–1281.
- [8] Ozawa, M.: *Heisenberg’s original derivation of the uncertainty principle and its universally valid reformulations*, Current Sci. **109** (2015) 2006–2016.
- [9] Dammeier, L., Schwonnek, R., Werner, R.F.: *Uncertainty relations for angular momentum*, New J. Phys. **17** (2015) 093046.
- [10] Abbott, A.A., Alzieu, P.-L., Hall, M.J.W., Branciard, C.: *Tight state-independent uncertainty relations for qubits*, Mathematics, **4** (2016) 8.
- [11] Werner, R.F.: *Uncertainty relations for general phase spaces*, Front. Phys. **11** (2016) 110305.
- [12] Heisenberg, W.: *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschr. Phys. **43** (1927) 172–198.
- [13] Robertson, H.: *The uncertainty principle*, Phys. Rev. **34** (1929) 163164.
- [14] Kraus, K.: *Complementary observables and uncertainty relations*, Phys. Rev. D **35** (1987) 3070–3075.
- [15] Maassen, H., Uffink, J.B.M.: *Generalized entropic uncertainty relations*, Phys. Rev. Lett. **60** (1988) 1103–1106.

- [16] Krishna, M., Parthasarathy, K.R.: *An Entropic Uncertainty Principle for Quantum Measurements*, Sankhya: Indian J. Stat. **64** (2002) 842–851.
- [17] Wehner, S., Winter, A.: *Entropic uncertainty relations — a survey*, New J. Phys. **12** (2010) 025009.
- [18] Abdelkhalek, K., Schwonnek, R., Maassen, H., Furrer, F., Duhme, J., Raynal, P., Englert, B-G., Werner, R.F.: *Optimality of entropic uncertainty relations*, Int. J. Quantum Inf. **13** (2015) 1550045.
- [19] Holevo, A.S.: *Statistical Structure of Quantum Theory*, Lecture Notes in Physics **m 67** (Springer, Berlin, 2001).
- [20] Busch, P., Grabowski, M., Lahti, P.: *Operational Quantum Physics* (Springer, Berlin, 1997).
- [21] Busch, P., Heinonen, T., Lahti, P.: *Heisenberg’s Uncertainty Principle*, Physics Reports **452** (2007) 155–176.
- [22] Busch, P., Lahti, P., Pellonpää, J.-P., Ylinen, K.: *Quantum Measurement* (Springer, Berlin, 2016).
- [23] Heinosaari, T., Wolf, M.M.: *Nondisturbing quantum measurements*, J. Math. Phys. **51** (2010) 092201.
- [24] Heinosaari, T., Miyadera, T.: *Universality of sequential quantum measurements*, Phys. Rev. **91** (2015) 022110.
- [25] Appleby, D.M.: *Error principle*, Internat. J. Theoret. Phys. **37** (1998) 25572572.
- [26] Appleby, D.M.: *Quantum Errors and Disturbances: Response to Busch, Lahti and Werner*, arXiv:1602.09002 [quant-ph].
- [27] Buscemi, F., Hall, M.J.W., Ozawa, M., Wilde, M.M.: *Noise and disturbance in quantum measurements: an information-theoretic approach*, Phys. Rev. Lett. **112** (2014) 050401.
- [28] Abbot, A.A., Branciard, C.: *Noise and disturbance of Qubit measurements: An information-theoretic characterisation*, arXiv:1607.00261.
- [29] Coles, P.J., Furrer, F.: *State-dependent approach to entropic measurement-disturbance relations*, Phys. Lett. A **379** (2015) 105–112.
- [30] Burnham, K.P., Anderson D.R.: *Model Selection and Multi-Model Inference*, 2nd edition (Springer, 2002).
- [31] Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, second edition (Wiley, 2006).
- [32] Ohya, M., Petz, D.: *Quantum entropy and its use* (Springer, Berlin, 1993).
- [33] Barchielli, A., Lupieri, G.: *Instruments and channels in quantum information theory*, Optics and Spectroscopy **99** (2005) 425–432.
- [34] Barchielli, A., Lupieri, G.: *Quantum measurements and entropic bounds on information transmission*, Quantum Inf. Comput. **6** (2006) 16–45.
- [35] Barchielli, A., Lupieri, G.: *Instruments and mutual entropies in quantum information*, Banach Center Publ. **73** (2006) 65–80.
- [36] Maccone, L.: *Entropic information-disturbance tradeoff*, Europhys. Lett. **77** (2007) 40002.
- [37] Davies, E.B.: *Quantum Theory of Open Systems* (Academic, London, 1976).
- [38] Holevo, A.S.: *Quantum Systems, Channels, Information* (de Gruiter, Berlin, 2012).
- [39] Heinosaari, T., Ziman, M.: *The mathematical language of quantum theory: From uncertainty to entanglement* (Cambridge University Press, Cambridge, 2012).

- [40] Topsøe, F.: *Basic concepts, identities and inequalities — the toolkit of Information Theory*, Entropy **3** (2001) 162–190.
- [41] Pedersen, G.K.: *Analysis now* (Springer-Verlag, New York, 1989).
- [42] Busch, P., Heinosaari, T.: *Approximate joint measurements of qubit observables*, Quantum Inf. Comp. **8** (2008) 797–818.
- [43] Heinosaari, T., Miyadera, T., Ziman, M.: *An invitation to quantum incompatibility*, J. Phys. A: Math. Theor. **49** (2016) 123001.
- [44] Busch, P., Heinosaari, T., Schultz, J., Stevens, N.: *Comparing the degrees of incompatibility inherent in probabilistic physical theories*, Europhys. Lett. **103** (2013) 10002.
- [45] Heinosaari, T., Schultz, J., Toigo, A., Ziman, M.: *Maximally incompatible quantum observables*, Phys. Lett. A **378** (2014) 1695–1699.
- [46] Keyl, M., Werner, R.F.: *Optimal cloning of pure states, testing single clones*, J. Math. Phys. **40** (1999) 3283–3299.
- [47] Werner, R.F.: *Optimal cloning of pure states*, Phys. Rev. A **58** (1998) 1827–1832.
- [48] P. Lahti, *Coexistence and Joint Measurability in Quantum Mechanics*, Int. J. Theor. Phys. **42** (2003) 893–906.
- [49] Carmeli, C., Heinosaari, T., Toigo, A.: *Sequential measurements of conjugate observables*, J. Phys. A: Math. Theor. **44** (2011) 285304.
- [50] Carmeli, C., Heinosaari, T., Toigo, A.: *Informationally complete joint measurements on finite quantum systems*,
- [51] Weyl, H.: *Symmetry* (Princeton University Press, 1952).
- [52] Wootters, W.K., Fields, D.B.: *Optimal state-determination by mutually unbiased measurements*, Ann. Physics **191** (1989) 363–381.
- [53] Durt, T., Englert, B.-G.; Bengtsson, I., Życzkowski, K.: *On mutually unbiased bases*, Int. J. Quantum Inf. **8** (2010) 535–640.
- [54] Bandyopadhyay, S., Boykin, P.O., Roychowdhury, V., Vatan, F.: *A new proof for the existence of mutually unbiased bases*, Algorithmica **34** (2002) 512–528.
- [55] Appleby, D.M.: *Properties of the extended Clifford group with applications to SIC-POVMs and MUBs*, arXiv:0909.5233.
- [56] Carmeli, C., Schultz, J., Toigo, A.: *Covariant mutually unbiased bases*, Rev. Math. Phys. **28** (2016) 1650009.
- [57] Lang, S.: *Algebra*, 3rd edition, Graduate Texts in Mathematics, No. 211 (Springer-Verlag, New York, 2002).
- [58] Carmeli, C., Heinosaari, T., Schultz, J., Toigo, A.: *Tasks and premises in quantum state determination*, J. Phys. A: Math. Theor. **47** (2014) 075302.
- [59] Busch, P., Lahti, P., Werner, R.F.: *Heisenberg uncertainty for qubit measurements*, Phys. Rev. A **89** (2014) 012129.
- [60] Heinosaari, T., Jivulescu, M.A., Reitzner, D., Ziman, M.: *Approximating incompatible von Neumann measurements simultaneously*, Phys. Rev. A **82** (2010) 032328.